Microeconomic Theory

Andreu Mas-Colell Michael D. Whinston

and

Jerry R. Green

New York Oxford OXFORD UNIVERSITY PRESS 1995

Contents

Preface

Microeconomic Theory is intended to serve as the text for a first-year graduate course in microeconomic theory.The original sources for much of the book's material are the lecture notes that we have provided over the years to students in the first-year microeconomic theory course at Harvard. Starting from these notes, we have tried to produce a text that covers in an accessible yet rigorous way the full range of topics taught in a typical first-year course.

The nonlexicographic ordering of our names deserves some explanation. The project was first planned and begun by the three of us in the spring of 1990. However, in February 1992, after early versions of most of the book's chapters had been drafted, Jerry Green was selected to serve as Provost of Harvard University, a position that forced him to suspend his involvement in the project. From this point in time until the manuscript's completion in June 1994, Andreu Mas-Colell and Michael Whinston assumed full responsibility for the project. With the conclusion of Jerry Green's service as Provost, the original three-person team was reunited for the review of galley and page proofs during the winter of 1994/1995.

The Organization of the Book

Microeconomic theory as a discipline begins by considering the behavior of individual agents and builds from this foundation to a theory of aggregate economic outcomes. *Microeconomic Theory* (the book) follows exactly this outline. It is divided into five parts. Part I covers individual decision making. It opens with a general treatment of individual choice and proceeds to develop the classical theories of consumer and producer behavior. It also provides an introduction to the theory of individual choice under uncertainty. Part II covers game theory, the extension of the theory of individual decision making to situations in which several decision makers interact. Part III initiates the investigation of market equilibria. It begins with an introduction to competitive equilibrium and the fundamental theorems of welfare economics in the context of the Marshallian partial equilibrium model. It then explores the possibilities for market failures in the presence of externalities, market power, and asymmetric information. Part IV substantially extends our previous study of competitive markets to the general equilibrium context. The positive and normative aspects of the theory are examined in detail, as are extensions of the theory to equilibrium under uncertainty and over time. Part V studies welfare economics. It discusses the possibilities for aggregation of individual preferences into social preferences both with and without interpersonal utility comparisons, as well as the implementation of social choices in the presence of incomplete information about agents' preferences. A Mathematical Appendix provides an introduction to most of the more advanced mathematics used in the book (e.g., concave/convex

functions, constrained optimization techniques, fixed point theorems, etc.) as well as references for further reading.

The Style of the Book

In choosing the content of *Microeconomic Theory* we have tried to err on the side of inclusion. Our aim has been to assure coverage of most topics that instructors in a first-year graduate microeconomic theory course might want to teach. An inevitable consequence of this choice is that the book covers more topics than any single first-year course can discuss adequately. (We certainly have never taught all of it in any one year.) Our hope is that the range of topics presented will allow instructors the freedom to emphasize those they find most important.

We have sought a style of presentation that is accessible, yet also rigorous. Wherever possible we give precise definitions and formal proofs of propositions. At the same time, we accompany this analysis with extensive verbal discussion as well as with numerous examples to illustrate key concepts. Where we have considered a proof or topic either too difficult or too peripheral we have put it into smaller type to allow students to skip over it easily in a first reading.

Each chapter offers many exercises, ranging from easy to hard [graded from A (easiest) to C (hardest)] to help students master the material. Some of these exercises also appear within the text of the chapters so that students can check their understanding along the way (almost all of these are level A exercises).

The mathematical prerequisites for use of the book are a basic knowledge of calculus, some familiarity with linear algebra (although the use of vectors and matrices is introduced gradually in Part I), and a grasp of the elementary aspects of probability. Students also will find helpful some familiarity with microeconomics at the level of an intermediate undergraduate course.

Teaching the Book

The material in this book may be taught in many different sequences. Typically we have taught Parts I–III in the Fall semester and Parts IV and V in the Spring (omitting some topics in each case). A very natural alternative to this sequence (one used in a number of departments that we know of) might instead teach Parts I and IV in the Fall, and Parts II, III, and V in the Spring.[1] The advantage of this alternative sequence is that the study of general equilibrium analysis more closely follows the study of individual behavior in competitive markets that is developed in Part I. The disadvantage, and the reason we have not used this sequence in our own course, is that this makes for a more abstract first semester; our students have seemed happy to have the change of pace offered by game theory, oligopoly, and asymmetric information after studying Part I.

The chapters have been written to be relatively self-contained. As a result, they can be shifted easily among the parts to accommodate many other course sequences. For example, we have often opted to teach game theory on an "as needed" basis,

---

1. Obviously, some adjustment needs to be made by programs that operate on a quarter system.

breaking it up into segments that are discussed right before they are used (e.g., Chapter 7, Chapter 8, and Sections 9.A–B before studying oligopoly, Sections 9.C–D before covering signaling). Some other possibilities include teaching the aggregation of preferences (Chapter 21) immediately after individual decision making and covering the principal-agent problem (Chapter 14), adverse selection, signaling, and screening (Chapter 13), and mechanism design (Chapter 23) together in a section of the course focusing on information economics.

In addition, even within each part, the sequence of topics can often be altered easily. For example, it has been common in many programs to teach the preference-based theory of consumer demand before teaching the revealed preference, or "choice-based," theory. Although we think there are good reasons to reverse this sequence as we have done in Part I,[2] we have made sure that the material on demand can be covered in this more traditional way as well.[3]

On Mathematical Notation

For the most part, our use of mathematical notation is standard. Perhaps the most important mathematical rule to keep straight regards matrix notation. Put simply, vectors are always treated mathematically as *column vectors*, even though they are often displayed within the written text as rows to conserve space. The transpose of the (column) vector x is denoted by x^{T}. When taking the inner product of two (column) vectors x and y, we write $x \cdot y$; it has the same meaning as $x^{\mathrm{T}}y$. This and other aspects of matrix notation are reviewed in greater detail in Section M.A of the Mathematical Appendix.

To help highlight definitions and propositions we have chosen to display them in a different typeface than is used elsewhere in the text. One perhaps unfortunate consequence of this choice is that mathematical symbols sometimes appear slightly differently there than in the rest of the text. With this warning, we hope that no confusion will result.

Summation symbols ($\sum$) are displayed in various ways throughout the text. Sometimes they are written as

$$\sum_{n=1}^{N}$$

(usually only in displayed equations), but often to conserve space they appear as $\sum_{n=1}^{N}$, and in the many cases in which no confusion exists about the upper and lower limit of the index in the summation, we typically write just $\sum_n$. A similar point applies to the product symbol $\prod$.

---

2. In particular, it is *much* easier to introduce and derive many properties of demand in the choice-based theory than it is using the preference-based approach; and the choice-based theory gives you *almost* all the properties of demand that follow from assuming the existence of rational preferences.

3. To do this, one introduces the basics of the consumer's problem using Sections 2.A–D and 3.A–D, discusses the properties of uncompensated and compensated demand functions, the indirect utility function, and the expenditure function using Sections 3.D–I and 2.E, and then studies revealed preference theory using Sections 2.F and 3.J (and Chapter 1 for a more general overview of the two approaches).

Also described below are the meanings we attach to a few mathematical symbols whose use is somewhat less uniform in the literature [in this list, $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$ are (column) vectors, while X and Y are sets]:

Symbol	*Meaning*
$x \geq y$	$x_n \geq y_n$ for all $n = 1, \ldots, N$.
$x \gg y$	$x_n > y_n$ for all $n = 1, \ldots, N$.
$X \subset Y$	*weak* set inclusion ($x \in X$ implies $x \in Y$).
$X \backslash Y$	The set $\{x\colon x \in X \text{ but } x \notin Y\}$.
$E_x[f(x, y)]$	The expected value of the function $f(\cdot)$ over realizations of the random variable x. (When the expectation is over all of the arguments of the function we simply write $E[f(x, y)]$.)

Acknowledgments

Many people have contributed to the development of this book. Dilip Abreu, Doug Bernheim, David Card, Prajit Dutta, Steve Goldman, John Panzar, and David Pearce all (bravely) test-taught a very early version of the manuscript during the 1991–92 academic year. Their comments at that early stage were instrumental in the refinement of the book into its current style, and led to many other substantive improvements in the text. Our colleagues (and in some cases former students) Luis Corchón, Simon Grant, Drew Fudenberg, Chiaki Hara, Sergiu Hart, Bengt Holmstrom, Eric Maskin, John Nachbar, Martin Osborne, Ben Polak, Ariel Rubinstein, and Martin Weitzman offered numerous helpful suggestions. The book would undoubtedly have been better still had we managed to incorporate all of their ideas.

Many generations of first-year Harvard graduate students have helped us with their questions, comments, and corrections. In addition, a number of current and former students have played a more formal role in the book's development, serving as research assistants in various capacities. Shira Lewin read the entire manuscript, finding errors in our proofs, suggesting improvements in exposition, and even (indeed, often) correcting our grammar. Chiaki Hara, Ilya Segal, and Steve Tadelis, with the assistance of Marc Nachman, have checked that the book's many exercises could be solved, and have suggested how they might be formulated properly when our first attempt to do so failed. Chiaki Hara and Steve Tadelis have also given us extensive comments and corrections on the text itself. Emily Mechner, Nick Palmer, Phil Panet, and Billy Pizer were members of a team of first-year students that read our early drafts in the summer of 1992 and offered very helpful suggestions on how we could convey the material better.

Betsy Carpenter and Claudia Napolilli provided expert secretarial support throughout the project, helping to type some chapter drafts, copying material on very tight deadlines, and providing their support in hundreds of other ways. Gloria Gerrig kept careful track of our ever-increasing expenditures.

Our editor at Oxford, Herb Addison, was instrumental in developing the test teaching program that so helped us in the book's early stages, and offered his support throughout the book's development. Leslie Phillips of Oxford took our expression of appreciation for the look of the Feynman Lectures, and turned it into a book [illegible]

staff at Keyword Publishing Services did an absolutely superb job editing and producing the book on a very tight schedule. Their complete professionalism has been deeply appreciated.

The influence of many other individuals on the book, although more indirect, has been no less important. Many of the exercises that appear in the book have been conceived over the years by others, both at Harvard and elsewhere. We have indicated our source for an exercise whenever we were aware of it. Good exercises are an enormously valuable resource. We thank the anonymous authors of many of the exercises that appear here.

The work of numerous scholars has contributed to our knowledge of the topics discussed in this book. Of necessity we have been able to provide references in each chapter to only a limited number of sources. Many interesting and important contributions have not been included. These usually can be found in the references of the works we do list; indeed, most chapters include at least one reference to a general survey of their topic.

We have also had the good fortune to teach the first-year graduate microeconomic theory course at Harvard in the years prior to writing this book with Ken Arrow, Dale Jorgenson, Steve Marglin, Eric Maskin, and Mike Spence, from whom we learned a great deal about microeconomics and its teaching.

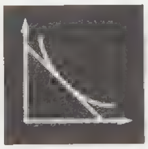
We also thank the NSF and Sloan Foundation for their support of our research over the years. In addition, the Center for Advanced Study in the Behavioral Sciences provided an ideal environment to Michael Whinston for completing the manuscript during the 1993/1994 academic year. The Universitat Pompeu Fabra also offered its hospitality to Andreu Mas-Colell at numerous points during the book's development.

Finally, we want to offer a special thanks to those who first excited us about the subject matter that appears here: Gerard Debreu, Leo Hurwicz, Roy Radner, Marcel Richter, and Hugo Sonnenschein (A.M.-C.); David Cass, Peter Diamond, Franklin Fisher, Sanford Grossman, and Eric Maskin (M.D.W.); Emmanuel Drandakis, Ron Jones, Lionel McKenzie, and Edward Zabel (J.R.G.).

A.M.-C., M.D.W., J.R.G.

Cambridge, MA
March 1995

PART ONE



Individual Decision Making

A distinctive feature of microeconomic theory is that it aims to model economic activity as an interaction of individual economic agents pursuing their private interests. It is therefore appropriate that we begin our study of microeconomic theory with an analysis of individual decision making.

Chapter 1 is short and preliminary. It consists of an introduction to the theory of individual decision making considered in an abstract setting. It introduces the decision maker and her choice problem, and it describes two related approaches to modeling her decisions. One, the *preference-based approach*, assumes that the decision maker has a preference relation over her set of possible choices that satisfies certain rationality axioms. The other, the *choice-based approach*, focuses directly on the decision maker's choice behavior, imposing consistency restrictions that parallel the rationality axioms of the preference-based approach.

The remaining chapters in Part One study individual decision making in explicitly economic contexts. It is common in microeconomics texts—and this text is no exception—to distinguish between two sets of agents in the economy: *individual consumers* and *firms*. Because individual consumers own and run firms and therefore ultimately determine a firm's actions, they are in a sense the more fundamental element of an economic model. Hence, we begin our review of the theory of economic decision making with an examination of the consumption side of the economy.

Chapters 2 and 3 study the behavior of consumers in a market economy. Chapter 2 begins by describing the consumer's decision problem and then introduces the concept of the consumer's *demand function*. We then proceed to investigate the implications for the demand function of several natural properties of consumer demand. This investigation constitutes an analysis of consumer behavior in the spirit of the choice-based approach introduced in Chapter 1.

In Chapter 3, we develop the classical preference-based approach to consumer demand. Topics such as utility maximization, expenditure minimization, duality, integrability, and the measurement of welfare changes are studied there. We also discuss the relation between this theory and the choice-based approach studied in Chapter 2.

In economic analysis, the aggregate behavior of consumers is often more important than the behavior of any single consumer. In Chapter 4, we analyze the

extent to which the properties of individual demand discussed in Chapters 2 and 3 also hold for aggregate consumer demand.

In Chapter 5, we study the behavior of the firm. We begin by posing the firm's decision problem, introducing its technological constraints and the assumption of profit maximization. A rich theory, paralleling that for consumer demand, emerges. In an important sense, however, this analysis constitutes a first step because it takes the objective of profit maximization as a maintained hypothesis. In the last section of the chapter, we comment on the circumstances under which profit maximization can be derived as the desired objective of the firm's owners.

Chapter 6 introduces risk and uncertainty into the theory of individual decision making. In most economic decision problems, an individual's or firm's choices do not result in perfectly certain outcomes. The theory of decision making under uncertainty developed in this chapter therefore has wide-ranging applications to economic problems, many of which we discuss later in the book.

CHAPTER

1

Preference and Choice

1.A Introduction

In this chapter, we begin our study of the theory of individual decision making by considering it in a completely abstract setting. The remaining chapters in Part I develop the analysis in the context of explicitly economic decisions.

The starting point for any individual decision problem is a *set of possible* (*mutually exclusive*) *alternatives* from which the individual must choose. In the discussion that follows, we denote this set of alternatives abstractly by X. For the moment, this set can be anything. For example, when an individual confronts a decision of what career path to follow, the alternatives in X might be: {go to law school, go to graduate school and study economics, go to business school, ..., become a rock star}. In Chapters 2 and 3, when we consider the consumer's decision problem, the elements of the set X are the possible consumption choices.

There are two distinct approaches to modeling individual choice behavior. The first, which we introduce in Section 1.B, treats the decision maker's tastes, as summarized in her *preference relation*, as the primitive characteristic of the individual. The theory is developed by first imposing rationality axioms on the decision maker's preferences and then analyzing the consequences of these preferences for her choice behavior (i.e., on decisions made). This preference-based approach is the more traditional of the two, and it is the one that we emphasize throughout the book.

The second approach, which we develop in Section 1.C, treats the individual's choice behavior as the primitive feature and proceeds by making assumptions directly concerning this behavior. A central assumption in this approach, the *weak axiom of revealed preference*, imposes an element of consistency on choice behavior, in a sense paralleling the rationality assumptions of the preference-based approach. This choice-based approach has several attractive features. It leaves room, in principle, for more general forms of individual behavior than is possible with the preference-based approach. It also makes assumptions about objects that are directly observable (choice behavior), rather than about things that are not (preferences). Perhaps most importantly, it makes clear that the theory of individual decision making need not be based on a process of introspection but can be given an entirely behavioral foundation.

Understanding the relationship between these two different approaches to modeling individual behavior is of considerable interest. Section 1.D investigates this question, examining first the implications of the preference-based approach for choice behavior and then the conditions under which choice behavior is compatible with the existence of underlying preferences. (This is an issue that also comes up in Chapters 2 and 3 for the more restricted setting of consumer demand.)

For an in-depth, advanced treatment of the material of this chapter, see Richter (1971).

1.B Preference Relations

In the preference-based approach, the objectives of the decision maker are summarized in a *preference relation*, which we denote by $\succsim$. Technically, $\succsim$ is a binary relation on the set of alternatives X, allowing the comparison of pairs of alternatives $x, y \in X$. We read $x \succsim y$ as "x is at least as good as y." From $\succsim$, we can derive two other important relations on X:

(i) The *strict preference* relation, $\succ$, defined by

$$x \succ y \Leftrightarrow x \succsim y \text{ but not } y \succsim x$$

and read "x is preferred to y."[1]

(ii) The *indifference* relation, $\sim$, defined by

$$x \sim y \Leftrightarrow x \succsim y \text{ and } y \succsim x$$

and read "x is indifferent to y."

In much of microeconomic theory, individual preferences are assumed to be *rational*. The hypothesis of rationality is embodied in two basic assumptions about the preference relation $\succsim$: *completeness* and *transitivity*.[2]

Definition 1.B.1: The preference relation $\succsim$ is *rational* if it possesses the following two properties:

(i) *Completeness:* for all $x, y \in X$, we have that $x \succsim y$ or $y \succsim x$ (or both).

(ii) *Transitivity:* For all $x, y, z \in X$, if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.

The assumption that $\succsim$ is complete says that the individual has a well-defined preference between any two possible alternatives. The strength of the completeness assumption should not be underestimated. Introspection quickly reveals how hard it is to evaluate alternatives that are far from the realm of common experience. It takes work and serious reflection to find out one's own preferences. The completeness axiom says that this task has taken place: our decision makers make only meditated choices.

Transitivity is also a strong assumption, and it goes to the heart of the concept of

1. The symbol $\Leftrightarrow$ is read as "if and only if." The literature sometimes speaks of $x \succsim y$ as "x is weakly preferred to y" and $x \succ y$ as "x is strictly preferred to y." We shall adhere to the terminology introduced above.

2. Note that there is no unified terminology in the literature; *weak order* and *complete preorder* are common alternatives to the term *rational preference relation*. Also, in some presentations, the assumption that $\succsim$ is *reflexive* (defined as $x \succsim x$ for all $x \in X$) is added to the completeness and transitivity assumptions. This property is, in fact, implied by completeness and so is redundant.

rationality. Transitivity implies that it is impossible to face the decision maker with a sequence of pairwise choices in which her preferences appear to cycle: for example, feeling that an apple is at least as good as a banana and that a banana is at least as good as an orange but then also preferring an orange over an apple. Like the completeness property, the transitivity assumption can be hard to satisfy when evaluating alternatives far from common experience. As compared to the completeness property, however, it is also more fundamental in the sense that substantial portions of economic theory would not survive if economic agents could not be assumed to have transitive preferences.

The assumption that the preference relation $\succsim$ is complete and transitive has implications for the strict preference and indifference relations $\succ$ and $\sim$. These are summarized in Proposition 1.B.1, whose proof we forgo. (After completing this section, try to establish these properties yourself in Exercises 1.B.1 and 1.B.2.)

Proposition 1.B.1: If $\succsim$ is rational then:

(i) $\succ$ is both *irreflexive* ($x \succ x$ never holds) and *transitive* (if $x \succ y$ and $y \succ z$, then $x \succ z$).

(ii) $\sim$ is *reflexive* ($x \sim x$ for all x), *transitive* (if $x \sim y$ and $y \sim z$, then $x \sim z$), and *symmetric* (if $x \sim y$, then $y \sim x$).

(iii) if $x \succ y \succsim z$, then $x \succ z$.

The irreflexivity of $\succ$ and the reflexivity and symmetry of $\sim$ are sensible properties for strict preference and indifference relations. A more important point in Proposition 1.B.1 is that rationality of $\succsim$ implies that both $\succ$ and $\sim$ are transitive. In addition, a transitive-like property also holds for $\succ$ when it is combined with an at-least-as-good-as relation, $\succsim$.

An individual's preferences may fail to satisfy the transitivity property for a number of reasons. One difficulty arises because of the problem of *just perceptible differences.* For example, if we ask an individual to choose between two very similar shades of gray for painting her room, she may be unable to tell the difference between the colors and will therefore be indifferent. Suppose now that we offer her a choice between the lighter of the two gray paints and a slightly lighter shade. She may again be unable to tell the difference. If we continue in this fashion, letting the paint colors get progressively lighter with each successive choice experiment, she may express indifference at each step. Yet, if we offer her a choice between the original (darkest) shade of gray and the final (almost white) color, she would be able to distinguish between the colors and is likely to prefer one of them. This, however, violates transitivity.

Another potential problem arises when the manner in which alternatives are presented matters for choice. This is known as the *framing* problem. Consider the following example, paraphrased from Kahneman and Tversky (1984):

> Imagine that you are about to purchase a stereo for 125 dollars and a calculator for 15 dollars. The salesman tells you that the calculator is on sale for 5 dollars less at the other branch of the store, located 20 minutes away. The stereo is the same price there. Would you make the trip to the other store?

It turns out that the fraction of respondents saying that they would travel to the other store for the 5 dollar discount is much higher than the fraction who say they would travel when the question is changed so that the 5 dollar saving is on the stereo. This is so even though the ultimate saving obtained by incurring the inconvenience of travel is the same in both

cases.[3] Indeed, we would expect indifference to be the response to the following question:

> Because of a stockout you must travel to the other store to get the two items, but you will receive 5 dollars off on either item as compensation. Do you care on which item this 5 dollar rebate is given?

If so, however, the individual violates transitivity. To see this, denote

x = Travel to the other store and get a 5 dollar discount on the calculator.
y = Travel to the other store and get a 5 dollar discount on the stereo.
z = Buy both items at the first store.

The first two choices say that $x \succ z$ and $z \succ y$, but the last choice reveals $x \sim y$. Many problems of framing arise when individuals are faced with choices between alternatives that have uncertain outcomes (the subject of Chapter 6). Kahneman and Tversky (1984) provide a number of other interesting examples.

At the same time, it is often the case that apparently intransitive behavior can be explained fruitfully as the result of the interaction of several more primitive rational (and thus transitive) preferences. Consider the following two examples

(i) A household formed by Mom (M), Dad (D), and Child (C) makes decisions by majority voting. The alternatives for Friday evening entertainment are attending an opera (O), a rock concert (R), or an ice-skating show (I). The three members of the household have the rational individual preferences: $O \succ_M R \succ_M I$, $I \succ_D O \succ_D R$, $R \succ_C I \succ_C O$, where $\succ_M$, $\succ_D$, $\succ_C$ are the transitive individual strict preference relations. Now imagine three majority-rule votes: O versus R, R versus I, and I versus O. The result of these votes (O will win the first, R the second, and I the third) will make the household's preferences $\succsim$ have the intransitive form: $O \succ R \succ I \succ O$. (The intransitivity illustrated in this example is known as the *Condorcet paradox*, and it is a central difficulty for the theory of group decision making. For further discussion, see Chapter 21.)

(ii) Intransitive decisions may also sometimes be viewed as a manifestation of a change of tastes. For example, a potential cigarette smoker may prefer smoking one cigarette a day to not smoking and may prefer not smoking to smoking heavily. But once she is smoking one cigarette a day, her tastes may change, and she may wish to increase the amount that she smokes. Formally, letting y be abstinence, x be smoking one cigarette a day, and z be heavy smoking, her initial situation is y, and her preferences in that initial situation are $x \succ y \succ z$. But once x is chosen over y and z, and there is a change of the individual's current situation from y to x, her tastes change to $z \succ x \succ y$. Thus, we apparently have an intransitivity: $z \succ x \succ z$. This *change-of-tastes* model has an important theoretical bearing on the analysis of addictive behavior. It also raises interesting issues related to commitment in decision making [see Schelling (1979)]. A rational decision maker will anticipate the induced change of tastes and will therefore attempt to tie her hand to her initial decision (Ulysses had himself tied to the mast when approaching the island of the Sirens). kontekstiz

It often happens that this change-of-tastes point of view gives us a well-structured way to think about *nonrational* decisions. See Elster (1979) for philosophical discussions of this and similar points.

Utility Functions

In economics, we often describe preference relations by means of a *utility function*. A utility function $u(x)$ assigns a numerical value to each element in X, ranking the

3. Kahneman and Tversky attribute this finding to individuals keeping "mental accounts" in which the savings are compared to the price of the item on which they are received.

elements of X in accordance with the individual's preferences. This is stated more precisely in Definition 1.B.2.

Definition 1.B.2: A function $u: X \to \mathbb{R}$ is a *utility function representing preference relation* $\succsim$ if, for all $x, y \in X$,

$$x \succsim y \Leftrightarrow u(x) \geq u(y).$$

Note that a utility function that represents a preference relation $\succsim$ is not unique. For any strictly increasing function $f: \mathbb{R} \to \mathbb{R}$, $v(x) = f(u(x))$ is a new utility function representing the same preferences as $u(\cdot)$; see Exercise 1.B.3. It is only the ranking of alternatives that matters. Properties of utility functions that are invariant for any strictly increasing transformation are called *ordinal*. *Cardinal* properties are those not preserved under all such transformations. Thus, the preference relation associated with a utility function is an ordinal property. On the other hand, the numerical values associated with the alternatives in X, and hence the magnitude of any differences in the utility measure between alternatives, are cardinal properties.

The ability to represent preferences by a utility function is closely linked to the assumption of rationality. In particular, we have the result shown in Proposition 1.B.2.

Proposition 1.B.2: A preference relation $\succsim$ can be represented by a utility function only if it is rational.

Proof: To prove this proposition, we show that if there is a utility function that represents preferences $\succsim$, then $\succsim$ must be complete and transitive.

Completeness. Because $u(\cdot)$ is a real-valued function defined on X, it must be that for any $x, y \in X$, either $u(x) \geq u(y)$ or $u(y) \geq u(x)$. But because $u(\cdot)$ is a utility function representing $\succsim$, this implies either that $x \succsim y$ or that $y \succsim x$ (recall Definition 1.B.2). Hence, $\succsim$ must be complete.

Transitivity. Suppose that $x \succsim y$ and $y \succsim z$. Because $u(\cdot)$ represents $\succsim$, we must have $u(x) \geq u(y)$ and $u(y) \geq u(z)$. Therefore, $u(x) \geq u(z)$. Because $u(\cdot)$ represents $\succsim$, this implies $x \succsim z$. Thus, we have shown that $x \succsim y$ and $y \succsim z$ imply $x \succsim z$, and so transitivity is established. ■

At the same time, one might wonder, can *any* rational preference relation $\succsim$ be described by some utility function? It turns out that, in general, the answer is no. An example where it is not possible to do so will be discussed in Section 3.G. One case in which we can always represent a rational preference relation with a utility function arises when X is finite (see Exercise 1.B.5). More interesting utility representation results (e.g., for sets of alternatives that are not finite) will be presented in later chapters.

1.C Choice Rules

In the second approach to the theory of decision making, choice behavior itself is taken to be the primitive object of the theory. Formally, choice behavior is represented by means of a *choice structure*. A choice structure $(\mathscr{B}, C(\cdot))$ consists of two ingredients:

(i) $\mathscr{B}$ is a family (a set) of nonempty subsets of X; that is, every element of $\mathscr{B}$ is a set $B \subset X$. By analogy with the consumer theory to be developed in Chapters 2 and 3, we call the elements $B \in \mathscr{B}$ *budget sets*. The budget sets in $\mathscr{B}$ should be thought of as an exhaustive listing of all the choice experiments that the institutionally, physically, or otherwise restricted social situation can conceivably pose to the decision maker. It need not, however, include all possible subsets of X. Indeed, in the case of consumer demand studied in later chapters, it will not.

(ii) $C(\cdot)$ is a *choice rule* (technically, it is a correspondence) that assigns a nonempty set of chosen elements $C(B) \subset B$ for every budget set $B \in \mathscr{B}$. When $C(B)$ contains a single element, that element is the individual's choice from among the alternatives in B. The set $C(B)$ may, however, contain more than one element. When it does, the elements of $C(B)$ are the alternatives in B that the decision maker *might* choose; that is, they are her *acceptable alternatives* in B. In this case, the set $C(B)$ can be thought of as containing those alternatives that we would actually see chosen if the decision maker were repeatedly to face the problem of choosing an alternative from set B.

Example 1.C.1: Suppose that $X = \{x, y, z\}$ and $\mathscr{B} = \{\{x, y\}, \{x, y, z\}\}$. One possible choice structure is $(\mathscr{B}, C_1(\cdot))$, where the choice rule $C_1(\cdot)$ is: $C_1(\{x, y\}) = \{x\}$ and $C_1(\{x, y, z\}) = \{x\}$. In this case, we see x chosen no matter what budget the decision maker faces.

Another possible choice structure is $(\mathscr{B}, C_2(\cdot))$, where the choice rule $C_2(\cdot)$ is: $C_2(\{x, y\}) = \{x\}$ and $C_2(\{x, y, z\}) = \{x, y\}$. In this case, we see x chosen whenever the decision maker faces budget $\{x, y\}$, but we may see either x or y chosen when she faces budget $\{x, y, z\}$. ■

When using choice structures to model individual behavior, we may want to impose some "reasonable" restrictions regarding an individual's choice behavior. An important assumption, the weak axiom of revealed preference [first suggested by Samuelson; see Chapter 5 in Samuelson (1947)], reflects the expectation that an individual's observed choices will display a certain amount of consistency. For example, if an individual chooses alternative x (and only that) when faced with a choice between x and y, we would be surprised to see her choose y when faced with a decision among x, y, and a third alterative z. The idea is that the choice of x when facing the alternatives $\{x, y\}$ reveals a proclivity for choosing x over y that we should expect to see reflected in the individual's behavior when faced with the alternatives $\{x, y, z\}$.[4]

The weak axiom is stated formally in Definition 1.C.1.

Definition 1.C.1: The choice structure $(\mathscr{B}, C(\cdot))$ satisfies the *weak axiom of revealed preference* if the following property holds:

If for some $B \in \mathscr{B}$ with $x, y \in B$ we have $x \in C(B)$, then for any $B' \in \mathscr{B}$ with $x, y \in B'$ and $y \in C(B')$, we must also have $x \in C(B')$.

In words, the weak axiom says that if x is ever chosen when y is available, then there can be no budget set containing both alternatives for which y is chosen and x is not.

4. This proclivity might reflect some underlying "preference" for x over y but might also arise in other ways. It could, for example, be the result of some evolutionary process.

Note how the assumption that choice behavior satisfies the weak axiom captures the consistency idea: If $C(\{x, y\}) = \{x\}$, then the weak axiom says that we cannot have $C(\{x, y, z\}) = \{y\}$.[5]

A somewhat simpler statement of the weak axiom can be obtained by defining a *revealed preference relation* $\succsim^*$ from the observed choice behavior in $C(\cdot)$.

Definition 1.C.2: Given a choice structure $(\mathscr{B}, C(\cdot))$ the *revealed preference relation* $\succsim^*$ is defined by

$$x \succsim^* y \Leftrightarrow \text{there is some } B \in \mathscr{B} \text{ such that } x, y \in B \text{ and } x \in C(B).$$

We read $x \succsim^* y$ as "x is revealed at least as good as y." Note that the revealed preference relation $\succsim^*$ need not be either complete or transitive. In particular, for any pair of alternatives x and y to be comparable, it is necessary that, for some $B \in \mathscr{B}$, we have $x, y \in B$ and either $x \in C(B)$ or $y \in C(B)$, or both.

We might also informally say that "x is revealed preferred to y" if there is some $B \in \mathscr{B}$ such that $x, y \in B$, $x \in C(B)$, and $y \notin C(B)$, that is, if x is ever chosen over y when both are feasible.

With this terminology, we can restate the weak axiom as follows: "*If x is revealed at least as good as y, then y cannot be revealed preferred to x.*"

Example 1.C.2: Do the two choice structures considered in Example 1.C.1 satisfy the weak axiom? Consider choice structure $(\mathscr{B}, C_1(\cdot))$. With this choice structure, we have $x \succsim^* y$ and $x \succsim^* z$, but there is no revealed preference relationship that can be inferred between y and z. This choice structure satisfies the weak axiom because y and z are never chosen.

Now consider choice structure $(\mathscr{B}, C_2(\cdot))$. Because $C_2(\{x, y, z\}) = \{x, y\}$, we have $y \succsim^* x$ (as well as $x \succsim^* y$, $x \succsim^* z$, and $y \succsim^* z$). But because $C_2(\{x, y\}) = \{x\}$, x is revealed preferred to y. Therefore, the choice structure $(\mathscr{B}, C_2)$ violates the weak axiom. ■

We should note that the weak axiom is not the only assumption concerning choice behavior that we may want to impose in any particular setting. For example, in the consumer demand setting discussed in Chapter 2, we impose further conditions that arise naturally in that context.

The weak axiom restricts choice behavior in a manner that parallels the use of the rationality assumption for preference relations. This raises a question: What is the precise relationship between the two approaches? In Section 1.D, we explore this matter.

1.D The Relationship between Preference Relations and Choice Rules

We now address two fundamental questions regarding the relationship between the two approaches discussed so far:

5. In fact, it says more: We must have $C(\{x, y, z\}) = \{x\}$, $=\{z\}$, or $=\{x, z\}$. You are asked to show this in Exercise 1.C.1. See also Exercise 1.C.2.

(i) If a decision maker has a rational preference ordering $\succsim$, do her decisions when facing choices from budget sets in $\mathscr{B}$ necessarily generate a choice structure that satisfies the weak axiom?

(ii) If an individual's choice behavior for a family of budget sets $\mathscr{B}$ is captured by a choice structure $(\mathscr{B}, C(\cdot))$ satisfying the weak axiom, is there necessarily a rational preference relation that is consistent with these choices?

As we shall see, the answers to these two questions are, respectively, "yes" and "maybe".

To answer the first question, suppose that an individual has a rational preference relation $\succsim$ on X. If this individual faces a nonempty subset of alternatives $B \subset X$, her preference-maximizing behavior is to choose any one of the elements in the set:

$$C^*(B, \succsim) = \{x \in B\colon x \succsim y \text{ for every } y \in B\}$$

The elements of set $C^*(B, \succsim)$ are the decision maker's most preferred alternatives in B. In principle, we could have $C^*(B, \succsim) = \varnothing$ for some B; but if X is finite, or if suitable (continuity) conditions hold, then $C^*(B, \succsim)$ will be nonempty.[6] From now on, we will consider only preferences $\succsim$ and families of budget sets $\mathscr{B}$ such that $C^*(B, \succsim)$ is nonempty for all $B \in \mathscr{B}$. We say that the rational preference relation $\succsim$ *generates* the choice structure $(\mathscr{B}, C^*(\cdot, \succsim))$.

The result in Proposition 1.D.1 tells us that any choice structure generated by rational preferences necessarily satisfies the weak axiom.

Proposition 1.D.1: Suppose that $\succsim$ is a rational preference relation. Then the choice structure generated by $\succsim$, $(\mathscr{B}, C^*(\cdot, \succsim))$, satisfies the weak axiom.

Proof: Suppose that for some $B \in \mathscr{B}$, we have $x, y \in B$ and $x \in C^*(B, \succsim)$. By the definition of $C^*(B, \succsim)$, this implies $x \succsim y$. To check whether the weak axiom holds, suppose that for some $B' \in \mathscr{B}$ with $x, y \in B'$, we have $y \in C^*(B', \succsim)$. This implies that $y \succsim z$ for all $z \in B'$. But we already know that $x \succsim y$. Hence, by transitivity, $x \succsim z$ for all $z \in B'$, and so $x \in C^*(B', \succsim)$. This is precisely the conclusion that the weak axiom demands. ■

Proposition 1.D.1 constitutes the "yes" answer to our first question. That is, if behavior is generated by rational preferences then it satisfies the consistency requirements embodied in the weak axiom.

In the other direction (from choice to preferences), the relationship is more subtle. To answer this second question, it is useful to begin with a definition.

Definition 1.D.1: Given a choice structure $(\mathscr{B}, C(\cdot))$, we say that the rational preference relation $\succsim$ *rationalizes* $C(\cdot)$ relative to $\mathscr{B}$ if

$$C(B) = C^*(B, \succsim)$$

for all $B \in \mathscr{B}$, that is, if $\succsim$ generates the choice structure $(\mathscr{B}, C(\cdot))$.

In words, the rational preference relation $\succsim$ rationalizes choice rule $C(\cdot)$ on $\mathscr{B}$ if the optimal choices generated by $\succsim$ (captured by $C^*(\cdot, \succsim)$) coincide with $C(\cdot)$ for

6. Exercise 1.D.2 asks you to establish the nonemptiness of $C^*(B, \succsim)$ for the case where X is finite. For general results, See Section M.F of the Mathematical Appendix and Section 3.C for a specific application.

all budget sets in $\mathscr{B}$. In a sense, preferences explain behavior; we can interpret the decision maker's choices as if she were a preference maximizer. Note that in general, there may be more than one rationalizing preference relation $\succsim$ for a given choice structure $(\mathscr{B}, C(\cdot))$ (see Exercise 1.D.1).

Proposition 1.D.1 implies that the weak axiom must be satisfied if there is to be a rationalizing preference relation. In particular, since $C^*(\cdot, \succsim)$ satisfies the weak axiom for any $\succsim$, only a choice rule that satisfies the weak axiom can be rationalized. It turns out, however, that the weak axiom is not sufficient to ensure the existence of a rationalizing preference relation.

Example 1.D.1: Suppose that $X = \{x, y, z\}$, $\mathscr{B} = \{\{x, y\}, \{y, z\}, \{x, z\}\}$, $C(\{x, y\}) = \{x\}$, $C(\{y, z\}) = \{y\}$, and $C(\{x, z\}) = \{z\}$. This choice structure satisfies the weak axiom (you should verify this). Nevertheless, we cannot have rationalizing preferences. To see this, note that to rationalize the choices under $\{x, y\}$ and $\{y, z\}$ it would be necessary for us to have $x \succ y$ and $y \succ z$. But, by transitivity, we would then have $x \succ z$, which contradicts the choice behavior under $\{x, z\}$. Therefore, there can be no rationalizing preference relation. ■

To understand Example 1.D.1, note that the more budget sets there are in $\mathscr{B}$, the more the weak axiom restricts choice behavior; there are simply more opportunities for the decision maker's choices to contradict one another. In Example 1.D.1, the set $\{x, y, z\}$ is not an element of $\mathscr{B}$. As it happens, this is crucial (see Exercises 1.D.3). As we now show in Proposition 1.D.2, if the family of budget sets $\mathscr{B}$ includes enough subsets of X, and if $(\mathscr{B}, C(\cdot))$ satisfies the weak axiom, then there exists a rational preference relation that rationalizes $C(\cdot)$ relative to $\mathscr{B}$ [this was first shown by Arrow (1959)].

Proposition 1.D.2: If $(\mathscr{B}, C(\cdot))$ is a choice structure such that

(i) the weak axiom is satisfied,
(ii) $\mathscr{B}$ includes all subsets of X of up to three elements,

then there is a rational preference relation $\succsim$ that rationalizes $C(\cdot)$ relative to $\mathscr{B}$; that is, $C(B) = C^*(B, \succsim)$ for all $B \in \mathscr{B}$. Furthermore, this rational preference relation is the *only* preference relation that does so.

Proof: The natural candidate for a rationalizing preference relation is the revealed preference relation $\succsim^*$. To prove the result, we must first show two things: (i) that $\succsim^*$ is a rational preference relation, and (ii) that $\succsim^*$ rationalizes $C(\cdot)$ on $\mathscr{B}$. We then argue, as point (iii), that $\succsim^*$ is the unique preference relation that does so.

(i) We first check that $\succsim^*$ is rational (i.e., that it satisfies completeness and transitivity).

Completeness By assumption (ii), $\{x, y\} \in \mathscr{B}$. Since either x or y must be an element of $C(\{x, y\})$, we must have $x \succsim^* y$, or $y \succsim^* x$, or both. Hence $\succsim^*$ is complete.

Transitivity Let $x \succsim^* y$ and $y \succsim^* z$. Consider the budget set $\{x, y, z\} \in \mathscr{B}$. It suffices to prove that $x \in C(\{x, y, z\})$, since this implies by the definition of $\succsim^*$ that $x \succsim^* z$. Because $C(\{x, y, z\}) \neq \varnothing$, at least one of the alternatives x, y, or z must be an element of $C(\{x, y, z\})$. Suppose that $y \in C(\{x, y, z\})$. Since $x \succsim^* y$, the weak axiom then yields $x \in C(\{x, y, z\})$, as we want. Suppose instead that $z \in C(\{x, y, z\})$; since $y \succsim^* z$, the weak axiom yields $y \in C(\{x, y, z\})$, and we are in the previous case.

(ii) We now show that $C(B) = C^*(B, \succsim^*)$ for all $B \in \mathscr{B}$; that is, the revealed preference

relation $\succsim^*$ inferred from $C(\cdot)$ actually generates $C(\cdot)$. Intuitively, this seems sensible. Formally, we show this in two steps. First, suppose that $x \in C(B)$. Then $x \succsim^* y$ for all $y \in B$; so we have $x \in C^*(B, \succsim^*)$. This means that $C(B) \subset C^*(B, \succsim^*)$. Next, suppose that $x \in C^*(B, \succsim^*)$. This implies that $x \succsim^* y$ for *all* $y \in B$; and so for each $y \in B$, there must exist some set $B_y \in \mathscr{B}$ such that $x, y \in B_y$ and $x \in C(B_y)$. Because $C(B) \neq \varnothing$, the weak axiom then implies that $x \in C(B)$. Hence, $C^*(B, \succsim^*) \subset C(B)$. Together, these inclusion relations imply that $C(B) = C^*(B, \succsim^*)$.

(iii) To establish uniqueness, simply note that because $\mathscr{B}$ includes all two-element subsets of X, the choice behavior in $C(\cdot)$ completely determines the pairwise preference relations over X of any rationalizing preference.

This completes the proof. ■

We can therefore conclude from Proposition 1.D.2 that for the special case in which choice is defined for all subsets of X, a theory based on choice satisfying the weak axiom is completely equivalent to a theory of decision making based on rational preferences. Unfortunately, this special case is too special for economics. For many situations of economic interest, such as the theory of consumer demand, choice is defined only for special kinds of budget sets. In these settings, the weak axiom does not exhaust the choice implications of rational preferences. We shall see in Section 3.J, however, that a strengthening of the weak axiom (which imposes more restrictions on choice behavior) provides a necessary and sufficient condition for behavior to be capable of being rationalized by preferences.

Definition 1.D.1 defines a rationalizing preference as one for which $C(B) = C^*(B, \succsim)$. An alternative notion of a rationalizing preference that appears in the literature requires only that $C(B) \subset C^*(B, \succsim)$; that is, $\succsim$ is said to rationalize $C(\cdot)$ on $\mathscr{B}$ if $C(B)$ is a subset of the most preferred choices generated by $\succsim$, $C^*(B, \succsim)$, for every budget $B \in \mathscr{B}$.

There are two reasons for the possible use of this alternative notion. The first is, in a sense, philosophical. We might want to allow the decision maker to resolve her indifference in some specific manner, rather than insisting that indifference means that anything might be picked. The view embodied in Definition 1.D.1 (and implicitly in the weak axiom as well) is that if she chooses in a specific manner then she is, de facto, not indifferent.

The second reason is empirical. If we are trying to determine from data whether an individual's choice is compatible with rational preference maximization, we will in practice have only a finite number of observations on the choices made from any given budget set B. If $C(B)$ represents the set of choices made with this limited set of observations, then because these limited observations might not reveal all the decision maker's preference maximizing choices, $C(B) \subset C^*(B, \succsim)$ is the natural requirement to impose for a preference relationship to rationalize observed choice data.

Two points are worth noting about the effects of using this alternative notion. First, it is a weaker requirement. Whenever we can find a preference relation that rationalizes choice in the sense of Definition 1.D.1, we have found one that does so in this other sense, too. Second, in the abstract setting studied here, to find a rationalizing preference relation in this latter sense is actually trivial: Preferences that have the individual indifferent among all elements of X will rationalize *any* choice behavior in this sense. When this alternative notion is used in the economics literature, there is always an insistence that the rationalizing preference relation should satisfy some additional properties that are natural restrictions for the specific economic context being studied.

REFERENCES

Arrow, K. (1959). Rational choice functions and orderings. *Econometrica* **26**: 121–27.
Elster, J. (1979). *Ulysses and the Sirens*. Cambridge, U.K.: Cambridge University Press.
Kahneman, D., and A. Tversky. (1984). Choices, values, and frames. *American Psychologist* **39**: 341–50.
Plott, C. R. (1973). Path independence, rationality and social choice. *Econometrica* **41**: 1075–91.
Richter, M. (1971). Rational choice. Chap. 2 in *Preferences, Utility and Demand*, edited by J. Chipman, L. Hurwicz, and H. Sonnenschein. New York: Harcourt Brace Jovanovich.
Samuelson, P. (1947). *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press.
Schelling, T. (1979). *Micromotives and Macrobehavior*. New York: Norton.
Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review* **34**: 275–86.

EXERCISES

1.B.1[B] Prove property (iii) of Proposition 1.B.1.

1.B.2[A] Prove properties (i) and (ii) of Proposition 1.B.1.

1.B.3[B] Show that if $f: \mathbb{R} \to \mathbb{R}$ is a strictly increasing function and $u: X \to \mathbb{R}$ is a utility function representing preference relation $\succsim$, then the function $v: X \to \mathbb{R}$ defined by $v(x) = f(u(x))$ is also a utility function representing preference relation $\succsim$.

1.B.4[A] Consider a rational preference relation $\succsim$. Show that if $u(x) = u(y)$ implies $x \sim y$ and if $u(x) > u(y)$ implies $x \succ y$, then $u(\cdot)$ is a utility function representing $\succsim$.

1.B.5[B] Show that if X is finite and $\succsim$ is a rational preference relation on X, then there is a utility function $u: X \to \mathbb{R}$ that represents $\succsim$. [*Hint*: Consider first the case in which the individual's ranking between any two elements of X is strict (i.e., there is never any indifference), and construct a utility function representing these preferences; then extend your argument to the general case.]

1.C.1[B] Consider the choice structure $(\mathscr{B}, C(\cdot))$ with $\mathscr{B} = (\{x, y\}, \{x, y, z\})$ and $C(\{x, y\}) = \{x\}$. Show that if $(\mathscr{B}, C(\cdot))$ satisfies the weak axiom, then we must have $C(\{x, y, z\}) = \{x\}$, $= \{z\}$, or $= \{x, z\}$.

1.C.2[B] Show that the weak axiom (Definition 1.C.1) is equivalent to the following property holding:

> Suppose that $B, B' \in \mathscr{B}$, that $x, y \in B$, and that $x, y \in B'$. Then if $x \in C(B)$ and $y \in C(B')$, we must have $\{x, y\} \subset C(B)$ and $\{x, y\} \subset C(B')$.

1.C.3[C] Suppose that choice structure $(\mathscr{B}, C(\cdot))$ satisfies the weak axiom. Consider the following two possible revealed preferred relations, $\succ^*$ and $\succ^{**}$:

$$x \succ^* y \Leftrightarrow \text{there is some } B \in \mathscr{B} \text{ such that } x, y \in B,\ x \in C(B), \text{ and } y \notin C(B)$$

$$x \succ^{**} y \Leftrightarrow x \succsim^* y \text{ but not } y \succsim^* x$$

where $\succsim^*$ is the revealed at-least-as-good-as relation defined in Definition 1.C.2.

(a) Show that $\succ^*$ and $\succ^{**}$ give the same relation over X; that is, for any $x, y \in X$, $x \succ^* y \Leftrightarrow x \succ^{**} y$. Is this still true if $(\mathscr{B}, C(\cdot))$ does not satisfy the weak axiom?

(b) Must $\succ^*$ be transitive?

(c) Show that if $\mathscr{B}$ includes all three-element subsets of X, then $\succ^*$ is transitive.

1.D.1[B] Give an example of a choice structure that can be rationalized by several preference relations. Note that if the family of budgets $\mathscr{B}$ includes all the two-element subsets of X, then there can be at most one rationalizing preference relation.

1.D.2[A] Show that if X is finite, then any rational preference relation generates a nonempty choice rule; that is, $C(B) \neq \varnothing$ for any $B \subset X$ with $B \neq \varnothing$.

1.D.3[B] Let $X = \{x, y, z\}$, and consider the choice structure $(\mathscr{B}, C(\cdot))$ with

$$\mathscr{B} = \{\{x, y\}, \{y, z\}, \{x, z\}, \{x, y, z\}\}$$

and $C(\{x, y\}) = \{x\}$, $C(\{y, z\}) = \{y\}$, and $C(\{x, z\}) = \{z\}$, as in Example 1.D.1. Show that $(\mathscr{B}, C(\cdot))$ must violate the weak axiom.

1.D.4[B] Show that a choice structure $(\mathscr{B}, C(\cdot))$ for which a rationalizing preference relation $\succsim$ exists satisfies the *path-invariance* property: For every pair $B_1, B_2 \in \mathscr{B}$ such that $B_1 \cup B_2 \in \mathscr{B}$ and $C(B_1) \cup C(B_2) \in \mathscr{B}$, we have $C(B_1 \cup B_2) = C(C(B_1) \cup C(B_2))$, that is, the decision problem can safely be subdivided. See Plott (1973) for further discussion.

1.D.5[C] Let $X = \{x, y, z\}$ and $\mathscr{B} = \{\{x, y\}, \{y, z), \{z, x\}\}$. Suppose that choice is now stochastic in the sense that, for every $B \in \mathscr{B}$, $C(B)$ is a frequency distribution over alternatives in B. For example, if $B = \{x, y\}$, we write $C(B) = (C_x(B), C_y(B))$, where $C_x(B)$ and $C_y(B)$ are nonnegative numbers with $C_x(B) + C_y(B) = 1$. We say that the stochastic choice function $C(\cdot)$ can be *rationalized by preferences* if we can find a probability distribution Pr over the six possible (strict) preference relations on X such that for every $B \in \mathscr{B}$, $C(B)$ is precisely the frequency of choices induced by Pr. For example, if $B = \{x, y\}$, then $C_x(B) = Pr(\{\succ : x \succ y\})$. This concept originates in Thurstone (1927), and it is of considerable econometric interest (indeed, it provides a theory for the error term in observable choice).

(a) Show that the stochastic choice function $C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\frac{1}{2}, \frac{1}{2})$ can be rationalized by preferences.

(b) Show that the stochastic choice function $C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\frac{1}{4}, \frac{3}{4})$ is not rationalizable by preferences.

(c) Determine the $0 < \alpha < 1$ at which $C(\{x, y\}) = C(\{y, z\}) = C(\{z, x\}) = (\alpha, 1 - \alpha)$ switches from rationalizable to nonrationalizable.

CHAPTER

2

Consumer Choice

2.A Introduction

The most fundamental decision unit of microeconomic theory is the *consumer*. In this chapter, we begin our study of consumer demand in the context of a market economy. By a *market economy*, we mean a setting in which the goods and services that the consumer may acquire are available for purchase at known prices (or, equivalently, are available for trade for other goods at known rates of exchange).

We begin, in Sections 2.B to 2.D, by describing the basic elements of the consumer's decision problem. In Section 2.B, we introduce the concept of *commodities*, the objects of choice for the consumer. Then, in Sections 2.C and 2.D, we consider the physical and economic constraints that limit the consumer's choices. The former are captured in the *consumption set*, which we discuss in Section 2.C; the latter are incorporated in Section 2.D into the consumer's *Walrasian budget set*.

The consumer's decision subject to these constraints is captured in the consumer's *Walrasian demand function*. In terms of the choice-based approach to individual decision making introduced in Section 1.C, the Walrasian demand function is the consumer's choice rule. We study this function and some of its basic properties in Section 2.E. Among them are what we call *comparative statics* properties: the ways in which consumer demand changes when economic constraints vary.

Finally, in Section 2.F, we consider the implications for the consumer's demand function of the *weak axiom of revealed preference*. The central conclusion we reach is that in the consumer demand setting, the weak axiom is essentially equivalent to the *compensated law of demand*, the postulate that prices and demanded quantities move in opposite directions for price changes that leave real wealth unchanged.

2.B Commodities

The decision problem faced by the consumer in a market economy is to choose consumption levels of the various goods and services that are available for purchase in the market. We call these goods and services *commodities*. For simplicity, we assume that the number of commodities is finite and equal to L (indexed by $\ell = 1, \dots, L$).

As a general matter, a *commodity vector* (or commodity *bundle*) is a list of amounts of the different commodities,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_L \end{bmatrix},$$

and can be viewed as a point in $\mathbb{R}^L$, the commodity *space*.[1]

We can use commodity vectors to represent an individual's consumption levels. The ℓth entry of the commodity vector stands for the amount of commodity ℓ consumed. We then refer to the vector as a *consumption vector* or *consumption bundle*.

Note that time (or, for that matter, location) can be built into the definition of a commodity. Rigorously, bread today and tomorrow should be viewed as distinct commodities. In a similar vein, when we deal with decisions under uncertainty in Chapter 6, viewing bread in different "states of nature" as different commodities can be most helpful.

Although commodities consumed at different times should be viewed rigorously as distinct commodities, in practice, economic models often involve some "time aggregation." Thus, one commodity might be "bread consumed in the month of February," even though, in principle, bread consumed at each instant in February should be distinguished. A primary reason for such time aggregation is that the economic data to which the model is being applied are aggregated in this way. The hope of the modeler is that the commodities being aggregated are sufficiently similar that little of economic interest is being lost.

We should also note that in some contexts it becomes convenient, and even necessary, to expand the set of commodities to include goods and services that may potentially be available for purchase but are not actually so and even some that may be available by means other than market exchange (say, the experience of "family togetherness"). For nearly all of what follows here, however, the narrow construction introduced in this section suffices.

2.C The Consumption Set

Consumption choices are typically limited by a number of physical constraints. The simplest example is when it may be impossible for the individual to consume a negative amount of a commodity such as bread or water.

Formally, the *consumption set* is a subset of the commodity space $\mathbb{R}^L$, denoted by $X \subset \mathbb{R}^L$, whose elements are the consumption bundles that the individual can conceivably consume given the physical constraints imposed by his environment.

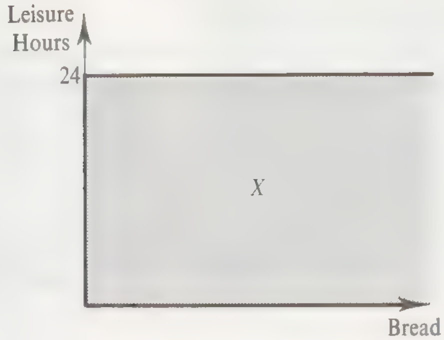
Consider the following four examples for the case in which $L = 2$:

(i) Figure 2.C.1 represents possible consumption levels of bread and leisure in a day. Both levels must be nonnegative and, in addition, the consumption of more than 24 hours of leisure in a day is impossible.

(ii) Figure 2.C.2 represents a situation in which the first good is perfectly divisible but the second is available only in nonnegative integer amounts.

(iii) Figure 2.C.3 captures the fact that it is impossible to eat bread at the same

1. Negative entries in commodity vectors will often represent debits or net outflows of goods. For example, in Chapter 5, the inputs of a firm are measured as negative numbers.



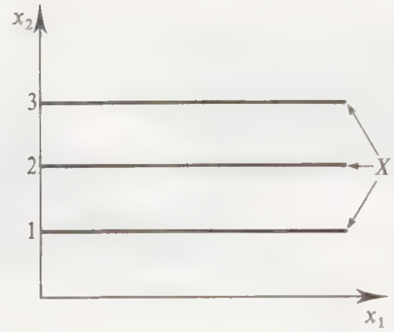
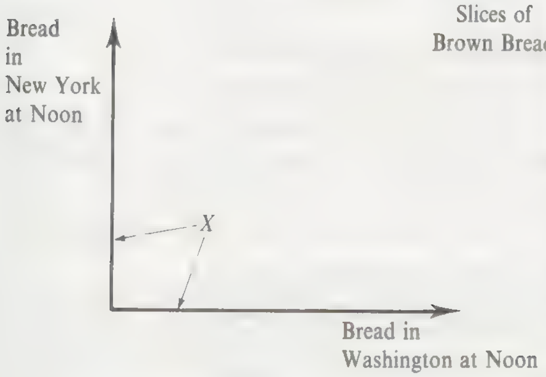


Figure 2.C.1 (left)
A consumption set.

Figure 2.C.2 (right)
A consumption set where good 2 must be consumed in integer amounts.



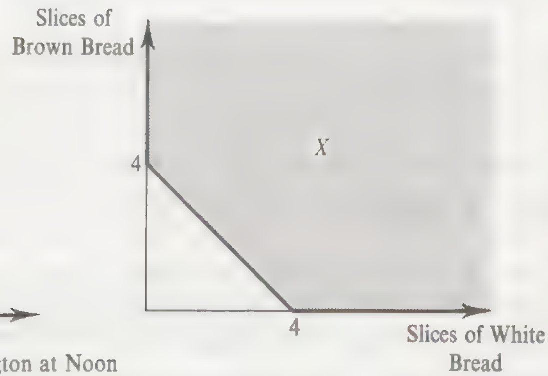


Figure 2.C.3 (left)
A consumption set where only one good can be consumed.

Figure 2.C.4 (right)
A consumption set reflecting survival needs.

instant in Washington and in New York. [This example is borrowed from Malinvaud (1978).]

(iv) Figure 2.C.4 represents a situation where the consumer requires a minimum of four slices of bread a day to survive and there are two types of bread, brown and white.

In the four examples, the constraints are physical in a very literal sense. But the constraints that we incorporate into the consumption set can also be institutional in nature. For example, a law requiring that no one work more than 16 hours a day would change the consumption set in Figure 2.C.1 to that in Figure 2.C.5.

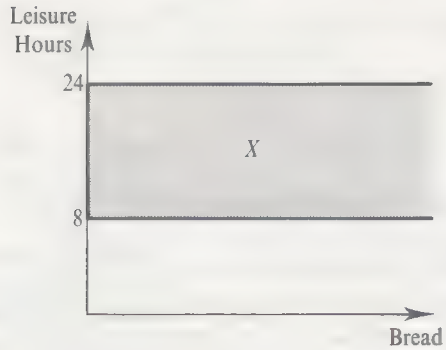
To keep things as straightforward as possible, we pursue our discussion adopting the simplest sort of consumption set:

$$X = \mathbb{R}^L_+ = \{x \in \mathbb{R}^L : x_\ell \geq 0 \text{ for } \ell = 1, \ldots, L\},$$

the set of all nonnegative bundles of commodities. It is represented in Figure 2.C.6. Whenever we consider any consumption set X other than $\mathbb{R}^L_+$, we shall be explicit about it.

One special feature of the set $\mathbb{R}^L_+$ is that it is *convex*. That is, if two consumption bundles x and x' are both elements of $\mathbb{R}^L_+$, then the bundle $x'' = \alpha x + (1 - \alpha)x'$ is also an element of $\mathbb{R}^L_+$ for any $\alpha \in [0, 1]$ (see Section M.G. of the Mathematical Appendix for the definition and properties of convex sets).[2] The consumption sets

2. Recall that $x'' = \alpha x + (1 - \alpha)x'$ is a vector whose ℓth entry is $x''_\ell = \alpha x_\ell + (1 - \alpha)x'_\ell$.



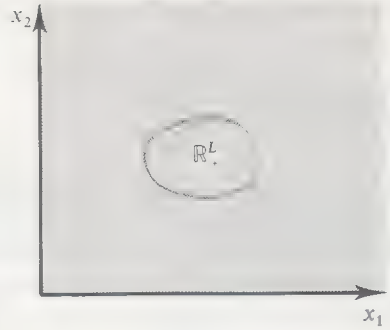


Figure 2.C.5 (left)
A consumption set reflecting a legal limit on the number of hours worked.

Figure 2.C.6 (right)
The consumption set $\mathbb{R}^L_+$.

in Figures 2.C.1, 2.C.4, 2.C.5, and 2.C.6 are convex sets; those in Figures 2.C.2 and 2.C.3 are not.

Much of the theory to be developed applies for general convex consumption sets as well as for $\mathbb{R}^L_+$. Some of the results, but not all, survive without the assumption of convexity.[3]

2.D Competitive Budgets

In addition to the physical constraints embodied in the consumption set, the consumer faces an important economic constraint: his consumption choice is limited to those commodity bundles that he can afford.

To formalize this constraint, we introduce two assumptions. First, we suppose that the L commodities are all traded in the market at dollar prices that are publicly quoted (this is the *principle of completeness*, or *universality, of markets*). Formally, these prices are represented by the *price vector*

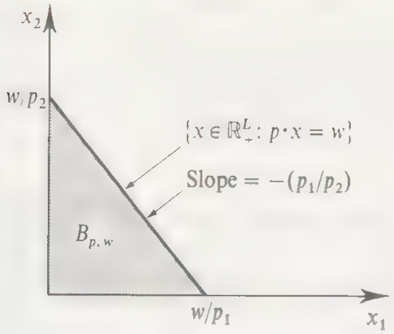
$$p = \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix} \in \mathbb{R}^L,$$

which gives the dollar cost for a unit of each of the L commodities. Observe that there is nothing that logically requires prices to be positive. A negative price simply means that a "buyer" is actually paid to consume the commodity (which is not illogical for commodities that are "bads," such as pollution). Nevertheless, for simplicity, here we always assume $p \gg 0$; that is, $p_\ell > 0$ for every ℓ.

Second, we assume that these prices are beyond the influence of the consumer. This is the so-called *price-taking assumption*. Loosely speaking, this assumption is likely to be valid when the consumer's demand for any commodity represents only a small fraction of the total demand for that good.

The affordability of a consumption bundle depends on two things: the market prices $p = (p_1, \ldots, p_L)$ and the consumer's wealth level (in dollars) w. The consumption

3. Note that commodity aggregation can help convexify the consumption set. In the example leading to Figure 2.C.3, the consumption set could reasonably be taken to be convex if the axes were instead measuring bread consumption over a period of a month.



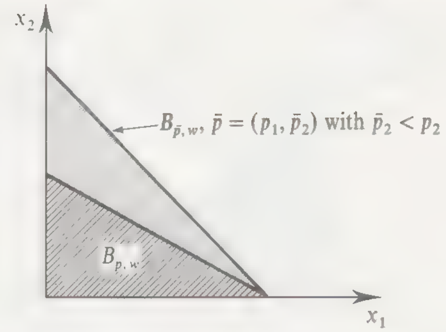


Figure 2.D.1 (left)
A Walrasian budget set.

Figure 2.D.2 (right)
The effect of a price change on the Walrasian budget set.

bundle $x \in \mathbb{R}^L_+$ is affordable if its total cost does not exceed the consumer's wealth level w, that is, if[4]

$$p \cdot x = p_1 x_1 + \cdots + p_L x_L \leq w.$$

This economic-affordability constraint, when combined with the requirement that x lie in the consumption set $\mathbb{R}^L_+$, implies that the set of feasible consumption bundles consists of the elements of the set $\{x \in \mathbb{R}^L_+ : p \cdot x \leq w\}$. This set is known as the *Walrasian* , *or competitive budget set* (after Léon Walras).

Definition 2.D.1: The *Walrasian, or competitive budget set* $B_{p,w} = \{x \in \mathbb{R}^L_+ : p \cdot x \leq w\}$ is the set of all feasible consumption bundles for the consumer who faces market prices p and has wealth w.

The *consumer's problem*, given prices p and wealth w, can thus be stated as follows: *Choose a consumption bundle x from $B_{p,w}$.*

A Walrasian budget set $B_{p,w}$ is depicted in Figure 2.D.1 for the case of $L = 2$. To focus on the case in which the consumer has a nondegenerate choice problem, we always assume $w > 0$ (otherwise the consumer can afford only $x = 0$).

The set $\{x \in \mathbb{R}^L_+ : p \cdot x = w\}$ is called the *budget hyperplane* (for the case $L = 2$, we call it the *budget line*). It determines the upper boundary of the budget set. As Figure 2.D.1 indicates, the slope of the budget line when $L = 2$, $-(p_1/p_2)$, captures the rate of exchange between the two commodities. If the price of commodity 2 decreases (with p_1 and w held fixed), say to $\bar{p}_2 < p_2$, the budget set grows larger because more consumption bundles are affordable, and the budget line becomes steeper. This change is shown in Figure 2.D.2.

Another way to see how the budget hyperplane reflects the relative terms of exchange between commodities comes from examining its geometric relation to the price vector p. The price vector p, drawn starting from any point $\bar{x}$ on the budget hyperplane, must be orthogonal (perpendicular) to any vector starting at $\bar{x}$ and lying

4. Often, this constraint is described in the literature as requiring that the cost of planned purchases not exceed the consumer's *income*. In either case, the idea is that the cost of purchases not exceed the consumer's available resources. We use the wealth terminology to emphasize that the consumer's actual problem may be intertemporal, with the commodities involving purchases over time, and the resource constraint being one of lifetime income (i.e., wealth) (see Exercise 2.D.1).

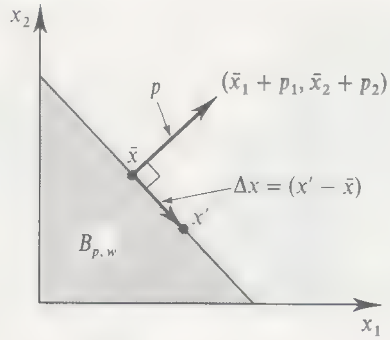


Figure 2.D.3 The geometric relationship between p and the budget hyperplane.

on the budget hyperplane. This is so because for any x' that itself lies on the budget hyperplane, we have $p \cdot x' = p \cdot \bar{x} = w$. Hence, $p \cdot \Delta x = 0$ for $\Delta x = (x' - \bar{x})$. Figure 2.D.3 depicts this geometric relationship for the case $L = 2$.[5]

The Walrasian budget set $B_{p,w}$ is a *convex* set: That is, if bundles x and x' are both elements of $B_{p,w}$, then the bundle $x'' = \alpha x + (1 - \alpha)x'$ is also. To see this, note first that because both x and x' are nonnegative, $x'' \in \mathbb{R}^L_+$. Second, since $p \cdot x \leq w$ and $p \cdot x' \leq w$, we have $p \cdot x'' = \alpha(p \cdot x) + (1 - \alpha)(p \cdot x') \leq w$. Thus, $x'' \in B_{p,w} = \{x \in \mathbb{R}^L_+ : p \cdot x \leq w\}$.

The convexity of $B_{p,w}$ plays a significant role in the development that follows. Note that the convexity of $B_{p,w}$ depends on the convexity of the consumption set $\mathbb{R}^L_+$. With a more general consumption set X, $B_{p,w}$ will be convex as long as X is. (See Exercise 2.D.3.)

Although Walrasian budget sets are of central theoretical interest, they are by no means the only type of budget set that a consumer might face in any actual situation. For example, a more realistic description of the market trade-off between a consumption good and leisure, involving taxes, subsidies, and several wage rates, is illustrated in Figure 2.D.4. In the figure, the price of the consumption good is 1, and the consumer earns wage rate s per hour for the first 8 hours of work and $s' > s$ for additional ("overtime") hours. He also faces a tax rate t

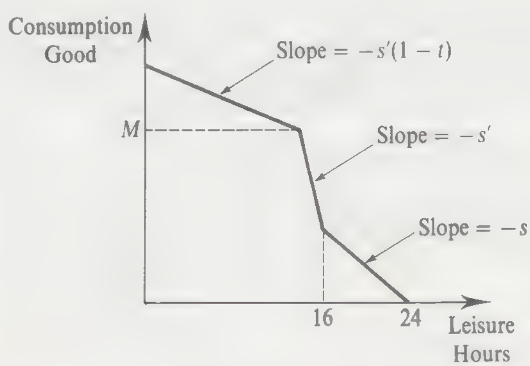


Figure 2.D.4 A more realistic description of the consumer's budget.

5. To draw the vector p starting from $\bar{x}$, we draw a vector from point $(\bar{x}_1, \bar{x}_2)$ to point $(\bar{x}_1 + p_1, \bar{x}_2 + p_2)$. Thus, when we draw the price vector in this diagram, we use the "units" on the axes to represent units of prices rather than goods.

per dollar on labor income earned above amount M. Note that the budget set in Figure 2.D.4 is not convex (you are asked to show this in Exercise 2.D.4). More complicated examples can readily be constructed and arise commonly in applied work. See Deaton and Muellbauer (1980) and Burtless and Hausmann (1975) for more illustrations of this sort.

2.E Demand Functions and Comparative Statics

The consumer's *Walrasian* (or *market*, or *ordinary*) *demand correspondence* $x(p, w)$ assigns a set of chosen consumption bundles for each price–wealth pair (p, w). In principle, this correspondence can be multivalued; that is, there may be more than one possible consumption vector assigned for a given price–wealth pair (p, w). When this is so, any $x \in x(p, w)$ might be chosen by the consumer when he faces price–wealth pair (p, w). When $x(p, w)$ is single-valued, we refer to it as a *demand function.*

Throughout this chapter, we maintain two assumptions regarding the Walrasian demand correspondence $x(p, w)$: That it is *homogeneous of degree zero* and that it satisfies *Walras' law.*

Definition 2.E.1: The Walrasian demand correspondence $x(p, w)$ is *homogeneous of degree zero* if $x(\alpha p, \alpha w) = x(p, w)$ for any p, w and $\alpha > 0$.

Homogeneity of degree zero says that if both prices and wealth change in the same proportion, then the individual's consumption choice does not change. To understand this property, note that a change in prices and wealth from (p, w) to $(\alpha p, \alpha w)$ leads to no change in the consumer's set of feasible consumption bundles; that is, $B_{p,w} = B_{\alpha p,\alpha w}$. Homogeneity of degree zero says that the individual's choice depends only on the set of feasible points.

Definition 2.E.2: The Walrasian demand correspondence $x(p, w)$ satisfies Walras' law if for every $p \gg 0$ and $w > 0$, we have $p \cdot x = w$ for all $x \in x(p, w)$.

Walras' law says that the consumer fully expends his wealth. Intuitively, this is a reasonable assumption to make as long as there is some good that is clearly desirable. Walras' law should be understood broadly: the consumer's budget may be an intertemporal one allowing for savings today to be used for purchases tomorrow. What Walras' law says is that the consumer fully expends his resources *over his lifetime.*

Exercise 2.E.1: Suppose $L = 3$, and consider the demand function $x(p, w)$ defined by

$$x_1(p, w) = \frac{p_2}{p_1 + p_2 + p_3} \frac{w}{p_1},$$

$$x_2(p, w) = \frac{p_3}{p_1 + p_2 + p_3} \frac{w}{p_2},$$

$$x_3(p, w) = \frac{\beta p_1}{p_1 + p_2 + p_3} \frac{w}{p_3}.$$

Does this demand function satisfy homogeneity of degree zero and Walras' law when $\beta = 1$? What about when $\beta \in (0,1)$?

In Chapter 3, where the consumer's demand $x(p, w)$ is derived from the maximization of preferences, these two properties (homogeneity of degree zero and satisfaction of Walras' law) hold under very general circumstances. In the rest of this chapter, however, we shall simply take them as assumptions about $x(p, w)$ and explore their consequences.

One convenient implication of $x(p, w)$ being homogeneous of degree zero can be noted immediately: Although $x(p, w)$ formally has $L + 1$ arguments, we can, with no loss of generality, fix (*normalize*) the level of one of the $L + 1$ independent variables at an arbitrary level. One common normalization is $p_\ell = 1$ for some ℓ. Another is $w = 1$.[6] Hence, the effective number of arguments in $x(p, w)$ is L.

For the remainder of this section, we assume that $x(p, w)$ is always single-valued. In this case, we can write the function $x(p, w)$ in terms of commodity-specific demand functions:

$$x(p, w) = \begin{bmatrix} x_1(p, w) \\ x_2(p, w) \\ \vdots \\ x_L(p, w) \end{bmatrix}.$$

When convenient, we also assume $x(p, w)$ to be continuous and differentiable.

The approach we take here and in Section 2.F can be viewed as an application of the choice-based framework developed in Chapter 1. The family of Walrasian budget sets is $\mathscr{B}^{\mathscr{W}} = \{B_{p,w}: p \gg 0, w > 0\}$. Moreover, by homogeneity of degree zero, $x(p, w)$ depends only on the budget set the consumer faces. Hence $(\mathscr{B}^{\mathscr{W}}, x(\cdot))$ is a choice structure, as defined in Section 1.C. Note that the choice structure $(\mathscr{B}^{\mathscr{W}}, x(\cdot))$ does not include all possible subsets of X (e.g., it does not include all two- and three-element subsets of X). This fact will be significant for the relationship between the choice-based and preference-based approaches to consumer demand.

Comparative Statics

We are often interested in analyzing how the consumer's choice varies with changes in his wealth and in prices. The examination of a change in outcome in response to a change in underlying economic parameters is known as *comparative statics* analysis.

Wealth effects

For fixed prices $\bar{p}$, the function of wealth $x(\bar{p}, w)$ is called the consumer's *Engel function*. Its image in $\mathbb{R}^L_+$, $E_{\bar{p}} = \{x(\bar{p}, w): w > 0\}$, is known as the *wealth expansion path*. Figure 2.E.1 depicts such an expansion path.

At any (p, w), the derivative $\partial x_\ell(p, w)/\partial w$ is known as the *wealth effect* for the ℓth good.[7]

6. We use normalizations extensively in Part IV.

7. It is also known as the *income effect* in the literature. Similarly, the wealth expansion path is sometimes referred to as an *income expansion path*.

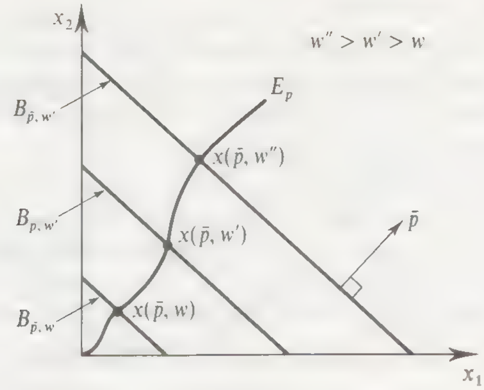


Figure 2.E.1
The wealth expansion path at prices $\bar{p}$.

A commodity ℓ is *normal* at (p, w) if $\partial x_\ell(p, w)/\partial w \geq 0$; that is, demand is nondecreasing in wealth. If commodity ℓ's wealth effect is instead negative, then it is called *inferior* at (p, w). If every commodity is normal at all (p, w), then we say that *demand is normal.*

The assumption of normal demand makes sense if commodities are large aggregates (e.g., food, shelter). But if they are very disaggregated (e.g., particular kinds of shoes), then because of substitution to higher-quality goods as wealth increases, goods that become inferior at some level of wealth may be the rule rather than the exception.

In matrix notation, the wealth effects are represented as follows:

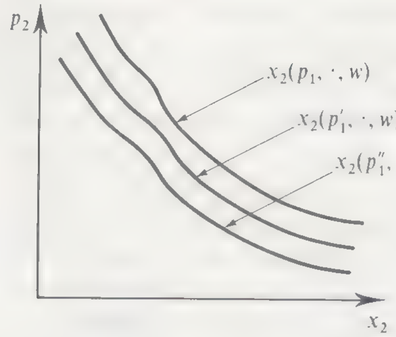
$$\frac{\partial x_\ell(p,w)}{\partial w} = \quad D_w x(p, w) = \begin{bmatrix} \dfrac{\partial x_1(p, w)}{\partial w} \\ \dfrac{\partial x_2(p, w)}{\partial w} \\ \vdots \\ \dfrac{\partial x_L(p, w)}{\partial w} \end{bmatrix} \in \mathbb{R}^L.$$

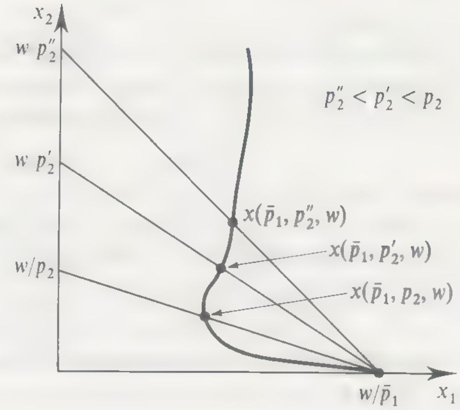
Price effects

We can also ask how consumption levels of the various commodities change as prices vary.

Consider first the case where $L = 2$, and suppose we keep wealth and price p_1 fixed. Figure 2.E.2 represents the demand function for good 2 as a function of its own price p_2 for various levels of the price of good 1, with wealth held constant at amount w. Note that, as is customary in economics, the price variable, which here is the independent variable, is measured on the vertical axis, and the quantity demanded, the dependent variable, is measured on the horizontal axis. Another useful representation of the consumers' demand at different prices is the locus of points demanded in $\mathbb{R}^2_+$ as we range over all possible values of p_2. This is known as an *offer curve.* An example is presented in Figure 2.E.3.

More generally, the derivative $\partial x_\ell(p, w)/\partial p_k$ is known as the *price effect of p_k,* the price of good k, *on the demand for good ℓ.* Although it may be natural to think that a fall in a good's price will lead the consumer to purchase more of it (as in





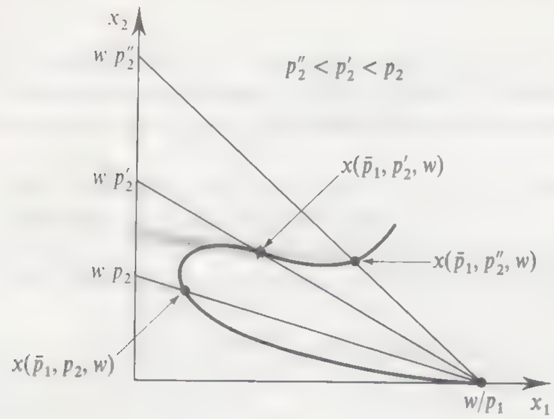


Figure 2.E.2 (top left)
The demand for good 2 as a function of its price (for various levels of p_1).

Figure 2.E.3 (top right)
An offer curve.

Figure 2.E.4 (bottom)
An offer curve where good 2 is inferior at $(\bar{p}, p_2', w)$.

Figure 2.E.3), the reverse situation is not an economic impossibility. Good ℓ is said to be a *Giffen good* at (p, w) if $\partial x_\ell(p, w)/\partial p_\ell > 0$. For the offer curve depicted in Figure 2.E.4, good 2 is a Giffen good at $(\bar{p}_1, p_2', w)$.

Low-quality goods may well be Giffen goods for consumers with low wealth levels. For example, imagine that a poor consumer initially is fulfilling much of his dietary requirements with potatoes because they are a low-cost way to avoid hunger. If the price of potatoes falls, he can then afford to buy other, more desirable foods that also keep him from being hungry. His consumption of potatoes may well fall as a result. Note that the mechanism that leads to potatoes being a Giffen good in this story involves a wealth consideration: When the price of potatoes falls, the consumer is effectively wealthier (he can afford to purchase more generally), and so he buys fewer potatoes. We will be investigating this interplay between price and wealth effects more extensively in the rest of this chapter and in Chapter 3.

The price effects are conveniently represented in matrix form as follows:

$\frac{\partial x_\ell(p, w)}{\partial p_k} =$

$$D_p x(p, w) = \begin{bmatrix} \frac{\partial x_1(p, w)}{\partial p_1} & \cdots & \frac{\partial x_1(p, w)}{\partial p_L} \\ & \ddots & \\ \frac{\partial x_L(p, w)}{\partial p_1} & \cdots & \frac{\partial x_L(p, w)}{\partial p_L} \end{bmatrix}.$$

Implications of homogeneity and Walras' law for price and wealth effects

Homogeneity and Walras' law imply certain restrictions on the comparative statics effects of consumer demand with respect to prices and wealth.

Consider, first, the implications of homogeneity of degree zero. We know that $x(\alpha p, \alpha w) - x(p, w) = 0$ for all $\alpha > 0$. Differentiating this expression with respect to α, and evaluating the derivative at $\alpha = 1$, we get the results shown in Proposition 2.E.1 (the result is also a special case of Euler's formula; see Section M.B of the Mathematical Appendix for details).

Proposition 2.E.1: If the Walrasian demand function $x(p, w)$ is homogeneous of degree zero, then for all p and w:

$$\sum_{k=1}^{L} \frac{\partial x_\ell(p, w)}{\partial p_k} p_k + \frac{\partial x_\ell(p, w)}{\partial w} w = 0 \text{ for } \ell = 1, \dots, L. \tag{2.E.1}$$

In matrix notation, this is expressed as

$$D_p x(p, w)p + D_w x(p, w) w = 0. \tag{2.E.2}$$

Thus, homogeneity of degree zero implies that the price and wealth derivatives of demand for any good ℓ, when weighted by these prices and wealth, sum to zero. Intuitively, this weighting arises because when we increase all prices and wealth proportionately, each of these variables changes in proportion to its initial level.

We can also restate equation (2.E.1) in terms of the *elasticities* of demand with respect to prices and wealth. These are defined, respectively, by

$$\varepsilon_{\ell k}(p, w) = \frac{\partial x_\ell(p, w)}{\partial p_k} \frac{p_k}{x_\ell(p, w)}$$

and

$$\varepsilon_{\ell w}(p, w) = \frac{\partial x_\ell(p, w)}{\partial w} \frac{w}{x_\ell(p, w)}.$$

These elasticities give the *percentage* change in demand for good ℓ per (marginal) percentage change in the price of good k or wealth; note that the expression for $\varepsilon_{\ell w}(\cdot, \cdot)$ can be read as $(\Delta x/x)/(\Delta w/w)$. Elasticities arise very frequently in applied work. Unlike the derivatives of demand, elasticities are independent of the units chosen for measuring commodities and therefore provide a unit-free way of capturing demand responsiveness.

Using elasticities, condition (2.E.1) takes the following form:

$$\sum_{k=1}^{L} \varepsilon_{\ell k}(p, w) + \varepsilon_{\ell w}(p, w) = 0 \quad \text{for } \ell = 1, \dots, L. \tag{2.E.3}$$

This formulation very directly expresses the comparative statics implication of homogeneity of degree zero: An equal percentage change in all prices and wealth leads to no change in demand.

Walras' law, on the other hand, has two implications for the price and wealth effects of demand. By Walras' law, we know that $p \cdot x(p, w) = w$ for all p and w. Differentiating this expression with respect to prices yields the first result, presented in Proposition 2.E.2.

Proposition 2.E.2: If the Walrasian demand function $x(p, w)$ satisfies Walras' law, then for all p and w:

$$\sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p, w)}{\partial p_k} + x_k(p, w) = 0 \quad \text{for } k = 1, \ldots, L, \tag{2.E.4}$$

or, written in matrix notion,[8]

$$p \cdot D_p x(p, w) + x(p, w)^{\mathrm{T}} = 0^{\mathrm{T}}. \tag{2.E.5}$$

Similarly, differentiating $p \cdot x(p, w) = w$ with respect to w, we get the second result, shown in Proposition 2.E.3.

Proposition 2.E.3: If the Walrasian demand function $x(p, w)$ satisfies Walras' law, then for all p and w:

$$\sum_{\ell=1}^{L} p_\ell \frac{\partial x_\ell(p, w)}{\partial w} = 1, \tag{2.E.6}$$

or, written in matrix notation,

$$p \cdot D_w x(p, w) = 1. \tag{2.E.7}$$

The conditions derived in Propositions 2.E.2 and 2.E.3 are sometimes called the properties of *Cournot* and *Engel aggregation*, respectively. They are simply the differential versions of two facts: That total expenditure cannot change in response to a change in prices and that total expenditure must change by an amount equal to any wealth change.

Exercise 2.E.2: Show that equations (2.E.4) and (2.E.6) lead to the following two elasticity formulas:

$$\sum_{\ell=1}^{L} b_\ell(p, w)\varepsilon_{\ell k}(p, w) + b_k(p, w) = 0,$$

$\varepsilon_{\ell k}(p,w) = \frac{\partial x_\ell(p,w)}{\partial p_k} \cdot \frac{p_k}{x_\ell}$

$= p_\ell x_\ell(p,w)/w \cdot \varepsilon_{\ell k}(p,w) + p_k x_k(p,w)/w$

and

$$\sum_{\ell=1}^{L} b_\ell(p, w)\varepsilon_{\ell w}(p, w) = 1,$$

$\varepsilon_{\ell w}(p,w) = \frac{\partial x_\ell(p,w)}{\partial w} \cdot \frac{w}{x_\ell}$

where $b_\ell(p, w) = p_\ell x_\ell(p, w)/w$ is the budget share of the consumer's expenditure on good ℓ given prices p and wealth w.

2.F The Weak Axiom of Revealed Preference and the Law of Demand

In this section, we study the implications of the weak axiom of revealed preference for consumer demand. Throughout the analysis, we continue to assume that $x(p, w)$ is single-valued, homogeneous of degree zero, and satisfies Walras' law.[9]

The weak axiom was already introduced in Section 1.C as a consistency axiom for the choice-based approach to decision theory. In this section, we explore its implications for the demand behavior of a consumer. In the preference-based approach to consumer behavior to be studied in Chapter 3, demand necessarily

8. Recall that 0^{T} means a row vector of zeros.
9. For generalizations to the case of multivalued choice, see Exercise 2.F.13.

satisfies the weak axiom. Thus, the results presented in Chapter 3, when compared with those in this section, will tell us how much more structure is imposed on consumer demand by the preference-based approach beyond what is implied by the weak axiom alone.[10]

In the context of Walrasian demand functions, the weak axiom takes the form stated in the Definition 2.F.1.

Definition 2.F.1: The Walrasian demand function $x(p, w)$ satisfies the *weak axiom of revealed preference* (the WA) if the following property holds for any two price–wealth situations (p, w) and (p', w'):

$$\text{If } p \cdot x(p', w') \leq w \quad \text{and} \quad x(p', w') \neq x(p, w), \quad \text{then} \quad p' \cdot x(p, w) > w'.$$

If you have already studied Chapter 1, you will recognize that this definition is precisely the specialization of the general statement of the weak axiom presented in Section 1.C to the context in which budget sets are Walrasian and $x(p, w)$ specifies a unique choice (see Exercise 2.F.1).

In the consumer demand setting, the idea behind the weak axiom can be put as follows: If $p \cdot x(p', w') \leq w$ and $x(p', w') \neq x(p, w)$, then we know that when facing prices p and wealth w, the consumer chose consumption bundle $x(p, w)$ even though bundle $x(p', w')$ was also affordable. We can interpret this choice as "revealing" a preference for $x(p, w)$ over $x(p', w')$. Now, we might reasonably expect the consumer to display some consistency in his demand behavior. In particular, given his revealed preference, we expect that he would choose $x(p, w)$ over $x(p', w')$ whenever they are both affordable. If so, bundle $x(p, w)$ must not be affordable at the price–wealth combination (p', w') at which the consumer chooses bundle $x(p', w')$. That is, as required by the weak axiom, we must have $p' \cdot x(p, w) > w'$.

The restriction on demand behavior imposed by the weak axiom when $L = 2$ is illustrated in Figure 2.F.1. Each diagram shows two budget sets $B_{p',w'}$ and $B_{p'',w''}$ and their corresponding choice $x(p', w')$ and $x(p'', w'')$. The weak axiom tells us that we cannot have both $p' \cdot x(p'', w'') \leq w'$ and $p'' \cdot x(p', w') \leq w''$. Panels (a) to (c) depict permissible situations, whereas demand in panels (d) and (e) violates the weak axiom.

Implications of the Weak Axiom

The weak axiom has significant implications for the effects of price changes on demand. We need to concentrate, however, on a special kind of price change.

As the discussion of Giffen goods in Section 2.E suggested, price changes affect the consumer in two ways. First, they alter the relative cost of different commodities. But, second, they also change the consumer's real wealth: An increase in the price of a commodity impoverishes the consumers of that commodity. To study the implications of the weak axiom, we need to isolate the first effect.

One way to accomplish this is to imagine a situation in which a change in prices is accompanied by a change in the consumer's wealth that makes his initial consumption bundle just affordable at the new prices. That is, if the consumer is originally facing prices p and wealth w and chooses consumption bundle $x(p, w)$, then

10. Or, stated more properly, beyond what is implied by the weak axiom in conjunction with homogeneity of degree zero and Walras' law.

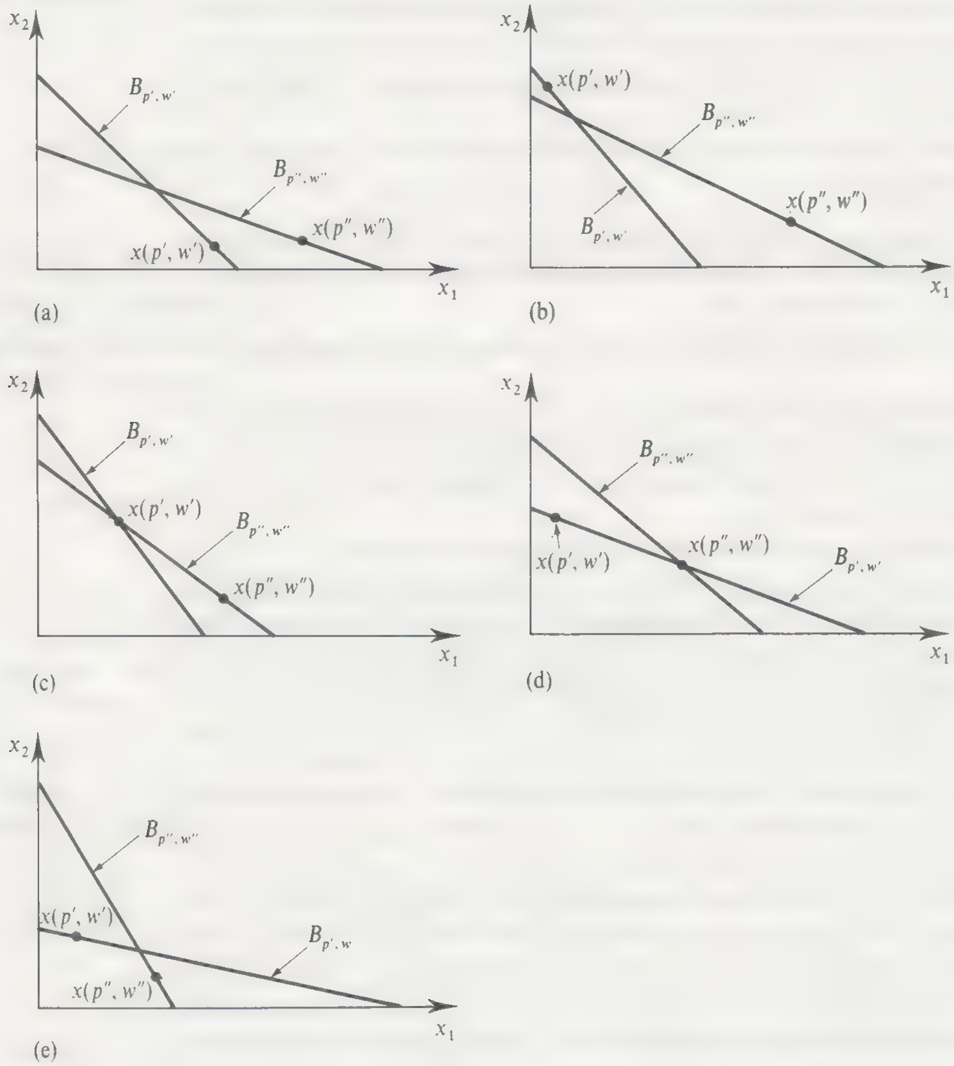


Figure 2.F.1
Demand in panels (a) to (c) satisfies the weak axiom; demand in panels (d) and (e) does not.

when prices change to p', we imagine that the consumer's wealth is adjusted to $w' = p' \cdot x(p, w)$. Thus, the wealth adjustment is $\Delta w = \Delta p \cdot x(p, w)$, where $\Delta p = (p' - p)$. This kind of wealth adjustment is known as *Slutsky wealth compensation.* Figure 2.F.2 shows the change in the budget set when a reduction in the price of good 1 from p_1 to p_1' is accompanied by Slutsky wealth compensation. Geometrically, the restriction is that the budget hyperplane corresponding to (p', w') goes through the vector $x(p, w)$.

We refer to price changes that are accompanied by such compensating wealth changes as *(Slutsky) compensated price changes.*

In Proposition 2.F.1, we show that the weak axiom can be equivalently stated in terms of the demand response to compensated price changes.

Proposition 2.F.1: Suppose that the Walrasian demand function $x(p, w)$ is homogeneous of degree zero and satisfies Walras' law. Then $x(p, w)$ satisfies the weak axiom if and only if the following property holds:

For any compensated price change from an initial situation (p, w) to a new price–wealth pair $(p', w') = (p', p' \cdot x(p, w))$, we have

$$(p' - p) \cdot [x(p', w') - x(p, w)] \leq 0, \tag{2.F.1}$$

with strict inequality whenever $x(p, w) \neq x(p', w')$.

Proof: (i) *The weak axiom implies inequality (2.F.1), with strict inequality if $x(p, w) \neq x(p', w')$.* The result is immediate if $x(p', w') = x(p, w)$, since then $(p' - p) \cdot [x(p', w') - x(p, w)] = 0$. So suppose that $x(p', w') \neq x(p, w)$. The left-hand side of inequality (2.F.1) can be written as

$$(p' - p) \cdot [x(p', w') - x(p, w)] = p' \cdot [x(p', w') - x(p, w)] - p \cdot [x(p', w') - x(p, w)]. \tag{2.F.2}$$

Consider the first term of (2.F.2). Because the change from p to p' is a compensated price change, we know that $p' \cdot x(p, w) = w'$. In addition, Walras' law tells us that $w' = p' \cdot x(p', w')$. Hence

$$p' \cdot [x(p', w') - x(p, w)] = 0. \tag{2.F.3}$$

Now consider the second term of (2.F.2). Because $p' \cdot x(p, w) = w'$, $x(p, w)$ is affordable under price–wealth situation (p', w'). The weak axiom therefore implies that $x(p', w')$ must *not* be affordable under price–wealth situation (p, w). Thus, we must have $p \cdot x(p', w') > w$. Since $p \cdot x(p, w) = w$ by Walras' law, this implies that

$$p \cdot [x(p', w') - x(p, w)] > 0 \tag{2.F.4}$$

Together, (2.F.2), (2.F.3) and (2.F.4) yield the result.

(ii) *The weak axiom is implied by (2.F.1) holding for all compensated price changes, with strict inequality if $x(p, w) \neq x(p', w')$.* The argument for this direction of the proof uses the following fact: The weak axiom holds if and only if it holds for all *compensated* price changes. That is, the weak axiom holds if, for any two price–wealth pairs (p, w) and (p', w'), we have $p' \cdot x(p, w) > w'$ whenever $p \cdot x(p', w') = w$ and $x(p', w') \neq x(p, w)$.

To prove the fact stated in the preceding paragraph, we argue that if the weak axiom is violated, then there must be a compensated price change for which it is violated. To see this, suppose that we have a violation of the weak axiom, that is, two price–wealth pairs (p', w') and (p'', w'') such that $x(p', w') \neq x(p'', w'')$, $p' \cdot x(p'', w'') \leq w'$, and $p'' \cdot x(p', w') \leq w''$. If one of these two weak inequalities holds with equality, then this is actually a compensated price change and we are done. So assume that, as shown in Figure 2.F.3, we have $p' \cdot x(p'', w'') < w'$ and $p'' \cdot x(p', w') < w''$.

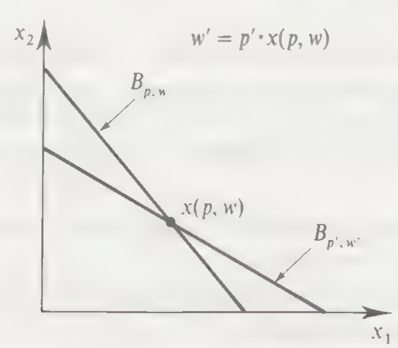


Figure 2.F.2
A compensated price change from (p, w) to (p', w').

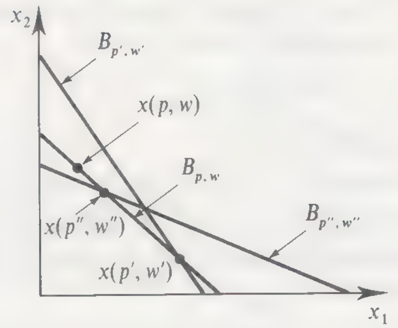


Figure 2.F.3
The weak axiom holds if and only if it holds for all compensated price changes.

Now choose the value of $\alpha \in (0,1)$ for which

$$(\alpha p' + (1-\alpha)p'')\cdot x(p', w') = (\alpha p' + (1-\alpha)p'')\cdot x(p'', w''),$$

and denote $p = \alpha p' + (1-\alpha)p''$ and $w = (\alpha p' + (1-\alpha)p'')\cdot x(p', w')$. This construction is illustrated in Figure 2.F.3. We then have

$$\begin{aligned}\alpha w' + (1-\alpha)w'' &> \alpha p'\cdot x(p', w') + (1-\alpha)p''\cdot x(p', w')\\ &= w\\ &= p\cdot x(p, w)\\ &= \alpha p'\cdot x(p, w) + (1-\alpha)p''\cdot x(p, w).\end{aligned}$$

Therefore, either $p'\cdot x(p, w) < w'$ or $p''\cdot x(p, w) < w''$. Suppose that the first possibility holds (the argument is identical if it is the second that holds). Then we have $x(p, w) \neq x(p', w')$, $p\cdot x(p', w') = w$, and $p'\cdot x(p, w) < w'$, which constitutes a violation of the weak axiom for the compensated price change from (p', w') to (p, w).

Once we know that in order to test for the weak axiom it suffices to consider only compensated price changes, the remaining reasoning is straightforward. If the weak axiom does not hold, there exists a compensated price change from some (p', w') to some (p, w) such that $x(p, w) \neq x(p', w')$, $p\cdot x(p', w') = w$, and $p'\cdot x(p, w) \leq w'$. But since $x(\cdot,\cdot)$ satisfies Walras' law, these two inequalities imply

$$p\cdot[x(p', w') - x(p, w)] = 0 \qquad \text{and} \qquad p'\cdot[x(p', w') - x(p, w)] \geq 0.$$

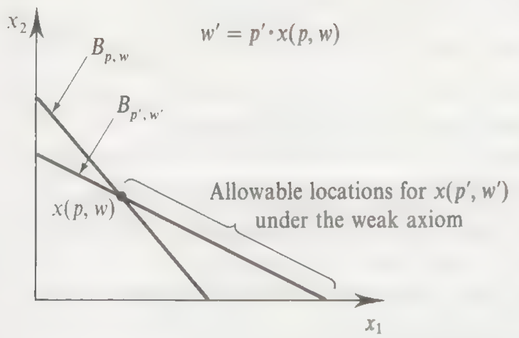
Hence, we would have

$$(p' - p)\cdot[x(p', w') - x(p, w)] \geq 0 \qquad \text{and} \qquad x(p, w) \neq x(p', w'),$$

which is a contradiction to (2.F.1) holding for all compensated price changes [and with strict inequality when $x(p, w) \neq x(p', w')$]. ■

The inequality (2.F.1) can be written in shorthand as $\Delta p\cdot\Delta x \leq 0$, where $\Delta p = (p' - p)$ and $\Delta x = [x(p', w') - x(p, w)]$. It can be interpreted as a form of the *law of demand*: *Demand and price move in opposite directions.* Proposition 2.F.1 tells us that the law of demand holds for *compensated* price changes. We therefore call it the *compensated law of demand.*

The simplest case involves the effect on demand for some good ℓ of a compensated change in its own price p_ℓ. When only this price changes, we have $\Delta p = (0, \ldots, 0, \Delta p_\ell, 0, \ldots, 0)$. Since $\Delta p\cdot\Delta x = \Delta p_\ell\,\Delta x_\ell$, Proposition 2.F.1 tells us that if $\Delta p_\ell > 0$, then we must have $\Delta x_\ell < 0$. The basic argument is illustrated in Figure 2.F.4. Starting at



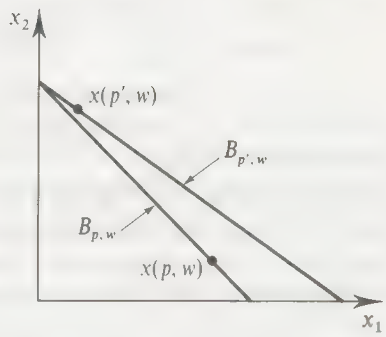


Figure 2.F.4 (left) Demand must be nonincreasing in own price for a compensated price change.

Figure 2.F.5 (right) Demand for good 1 can fall when its price decreases for an uncompensated price change.

(p, w), a compensated decrease in the price of good 1 rotates the budget line through $x(p, w)$. The WA allows moves of demand only in the direction that increases the demand of good 1.

Figure 2.F.5 should persuade you that the WA (or, for that matter, the preference maximization assumption discussed in Chapter 3) is not sufficient to yield the law of demand for price changes that are *not* compensated. In the figure, the price change from p to p' is obtained by a decrease in the price of good 1, but the weak axiom imposes no restriction on where we place the new consumption bundle; as drawn, the demand for good 1 falls.

When consumer demand $x(p, w)$ is a differentiable function of prices and wealth, Proposition 2.F.1 has a differential implication that is of great importance. Consider, starting at a given price–wealth pair (p, w), a differential change in prices dp. Imagine that we make this a compensated price change by giving the consumer compensation of $dw = x(p, w) \cdot dp$ [this is just the differential analog of $\Delta w = x(p, w) \cdot \Delta p$]. Proposition 2.F.1 tells us that

$$dp \cdot dx \leq 0. \tag{2.F.5}$$

Now, using the chain rule, the differential change in demand induced by this compensated price change can be written as

$$dx = D_p x(p, w)\, dp + D_w x(p, w)\, dw. \tag{2.F.6}$$

Hence

$$dx = D_p x(p, w)\, dp + D_w x(p, w)\, [x(p, w) \cdot dp] \tag{2.F.7}$$

or equivalently

$$dx = [D_p x(p, w) + D_w x(p, w) x(p, w)^{\mathrm{T}}]\, dp. \tag{2.F.8}$$

Finally, substituting (2.F.8) into (2.F.5) we conclude that for any possible differential price change dp, we have

$$dp \cdot [D_p x(p, w) + D_w x(p, w) x(p, w)^{\mathrm{T}}]\, dp \leq 0. \tag{2.F.9}$$

The expression in square brackets in condition (2.F.9) is an $L \times L$ matrix, which we denote by $S(p, w)$. Formally

$$S(p, w) = \begin{bmatrix} s_{11}(p, w) & \cdots & s_{1L}(p, w) \\ \vdots & \ddots & \vdots \\ s_{L1}(p, w) & \cdots & s_{LL}(p, w) \end{bmatrix},$$

where the (ℓ, k)th entry is

$$s_{\ell k}(p, w) = \frac{\partial x_\ell(p, w)}{\partial p_k} + \frac{\partial x_\ell(p, w)}{\partial w} x_k(p, w). \tag{2.F.10}$$

The matrix $S(p, w)$ is known as the *substitution*, or *Slutsky*, matrix, and its elements are known as *substitution effects*.

The "substitution" terminology is apt because the term $s_{\ell k}(p, w)$ measures the differential change in the consumption of commodity ℓ (i.e., the substitution to or from other commodities) due to a differential change in the price of commodity k when wealth is adjusted so that the consumer can still just afford his original consumption bundle (i.e., due solely to a change in relative prices). To see this, note that the change in demand for good ℓ if wealth is left unchanged is $(\partial x_l(p, w)/\partial p_k)\, dp_k$. For the consumer to be able to "just afford" his original consumption bundle, his wealth must vary by the amount $x_k(p, w)\, dp_k$. The effect of this wealth change on the demand for good ℓ is then $(\partial x_\ell(p, w)/\partial w)\,[x_k(p, w)\, dp_k]$. The sum of these two effects is therefore exactly $s_{\ell k}(p, w)\, dp_k$.

We summarize the derivation in equations (2.F.5) to (2.F.10) in Proposition 2.F.2.

Proposition 2.F.2: If a differentiable Walrasian demand function $x(p, w)$ satisfies Walras' law, homogeneity of degree zero, and the weak axiom, then at any (p, w), the Slutsky matrix $S(p, w)$ satisfies $v \cdot S(p, w) v \leq 0$ for any $v \in \mathbb{R}^L$.

A matrix satisfying the property in Proposition 2.F.2 is called *negative semidefinite* (it is negative *definite* if the inequality is strict for all $v \neq 0$). See Section M.D of the Mathematical Appendix for more on these matrices.

Note that $S(p, w)$ being negative semidefinite implies that $s_{\ell\ell}(p, w) \leq 0$: That is, *the substitution effect of good ℓ with respect to its own price is always nonpositive.*

An interesting implication of $s_{\ell\ell}(p, w) \leq 0$ is that a good can be a Giffen good at (p, w) only if it is inferior. In particular, since

$$s_{\ell\ell}(p, w) = \partial x_\ell(p, w)/\partial p_\ell + [\partial x_\ell(p, w)/\partial w]\, x_\ell(p, w) \leq 0,$$

if $\partial x_\ell(p, w)/\partial p_\ell > 0$, we must have $\partial x_\ell(p, w)/\partial w < 0$.

For later reference, we note that Proposition 2.F.2 does *not* imply, in general, that the matrix $S(p, w)$ is symmetric.[11] For $L = 2$, $S(p, w)$ is necessarily symmetric (you are asked to show this in Exercise 2.F.11). When $L > 2$, however, $S(p, w)$ need not be symmetric under the assumptions made so far (homogeneity of degree zero, Walras' law, and the weak axiom). See Exercises 2.F.10 and 2.F.15 for examples. In Chapter 3 (Section 3.H), we shall see that the symmetry of $S(p, w)$ is intimately connected with the possibility of generating demand from the maximization of rational preferences.

Exploiting further the properties of homogeneity of degree zero and Walras' law, we can say a bit more about the substitution matrix $S(p, w)$.

11. A matter of terminology: It is common in the mathematical literature that "definite" matrices are assumed to be symmetric. Rigorously speaking, if no symmetry is implied, the matrix would be called "quasidefinite." To simplify terminology, we use "definite" without any supposition about symmetry; if a matrix is symmetric, we say so explicitly. (See Exercise 2.F.9.)

Proposition 2.F.3: Suppose that the Walrasian demand function $x(p, w)$ is differentiable, homogeneous of degree zero, and satisfies Walras' law. Then $p \cdot S(p, w) = 0$ and $S(p, w)p = 0$ for any (p, w).

Exercise 2.F.7: Prove Proposition 2.F.3. [*Hint*: Use Propositions 2.E.1 to 2.E.3.]

It follows from Proposition 2.F.3 that the matrix $S(p, w)$ is always singular (i.e., it has rank less than L), and so the negative semidefiniteness of $S(p, w)$ established in Proposition 2.F.2 cannot be extended to negative definiteness (e.g., see Exercise 2.F.17).

Proposition 2.F.2 establishes negative semidefiniteness of $S(p, w)$ as a necessary implication of the weak axiom. One might wonder: Is this property sufficient to imply the WA [so that negative semidefiniteness of $S(p, w)$ is actually equivalent to the WA]? That is, if we have a demand function $x(p, w)$ that satisfies Walras' law, homogeneity of degree zero and has a negative semidefinite substitution matrix, must it satisfy the weak axiom? The answer is *almost, but not quite*. Exercise 2.F.16 provides an example of a demand function with a negative semidefinite substitution matrix that violates the WA. The sufficient condition is that $v \cdot S(p, w)v < 0$ whenever $v \neq \alpha p$ for any scalar α; that is, $S(p, w)$ must be negative definite for all vectors other than those that are proportional to p. This result is due to Samuelson [see Samuelson (1947) or Kihlstrom, Mas-Colell, and Sonnenschein (1976) for an advanced treatment]. The gap between the necessary and sufficient conditions is of the same nature as the gap between the necessary and the sufficient second-order conditions for the minimization of a function.

Finally, how would a theory of consumer demand that is based solely on the assumptions of homogeneity of degree zero, Walras' law, and the consistency requirement embodied in the weak action compare with one based on rational preference maximization?

Based on Chapter 1, you might hope that Proposition 1.D.2 implies that the two are equivalent. But we cannot appeal to that proposition here because the family of Walrasian budgets does not include every possible budget; in particular, it does not include all the budgets formed by only two- or three-commodity bundles.

In fact, the two theories are not equivalent. For Walrasian demand functions, the theory derived from the weak axiom is weaker than the theory derived from rational preferences, in the sense of implying fewer restrictions. This is shown formally in Chapter 3, where we demonstrate that if demand is generated from preferences, or is capable of being so generated, then it must have a symmetric Slutsky matrix at all (p, w). But for the moment, Example 2.F.1, due originally to Hicks (1956), may be persuasive enough.

Example 2.F.1: In a three-commodity world, consider the three budget sets determined by the price vectors $p^1 = (2, 1, 2)$, $p^2 = (2, 2, 1)$, $p^3 = (1, 2, 2)$ and wealth $= 8$ (the same for the three budgets). Suppose that the respective (unique) choices are $x^1 = (1, 2, 2)$, $x^2 = (2, 1, 2)$, $x^3 = (2, 2, 1)$. In Exercise 2.F.2, you are asked to verify that any two pairs of choices satisfy the WA but that x^3 is revealed preferred to x^2, x^2 is revealed preferred to x^1, and x^1 is revealed preferred to x^3. This situation is incompatible with the existence of underlying rational preferences (transitivity would be violated).

The reason this example is only *persuasive* and does not quite settle the question is that demand has been defined only for the three given budgets, therefore, we cannot be sure that it satisfies the requirements of the WA for all possible competitive budgets. To clinch the matter we refer to Chapter 3. ■

In summary, there are three primary conclusions to be drawn from Section 2.F:

(i) The consistency requirement embodied in the weak axiom (combined with the homogeneity of degree zero and Walras' law) is equivalent to the compensated law of demand.

(ii) The compensated law of demand, in turn, implies negative semidefiniteness of the substitution matrix $S(p, w)$.

(iii) These assumptions do *not* imply symmetry of $S(p, w)$, except in the case where $L = 2$.

REFERENCES

Burtless, G., and J. A. Hausman (1978). The effects of taxation on labor supply: Evaluating the Gary negative income tax experiment. *Journal of Political Economy* **86**: 1103–30.

Deaton, A., and J. Muellbauer (1980). *Economics and Consumer Behavior.* Cambridge, U.K.: Cambridge University Press.

Hicks, J. (1956). *A Revision of Demand Theory.* Oxford: Oxford University Press.

Kihlstrom, R., A. Mas-Colell, and H. Sonnenschein (1976). The demand theory of the weak axiom of revealed preference. *Econometrica* **44**: 971–78.

Malinvaud, E. (1978). *Lectures on Microeconomic Theory.* New York: Elsevier.

Samuelson, P. (1947). *Foundations of Economic Analysis.* Cambridge, Mass.: Harvard University Press.

EXERCISES

2.D.1^{A} A consumer lives for two periods, denoted 1 and 2, and consumes a single consumption good in each period. His wealth when born is $w > 0$. What is his (lifetime) Walrasian budget set?

2.D.2^{A} A consumer consumes one consumption good x and hours of leisure h. The price of the consumption good is p, and the consumer can work at a wage rate of $s = 1$. What is the consumer's Walrasian budget set?

2.D.3^{B} Consider an extension of the Walrasian budget set to an arbitrary consumption set X: $B_{p,w} = \{x \in X : p \cdot x \leq w\}$. Assume $(p, w) \gg 0$.

(a) If X is the set depicted in Figure 2.C.3, would $B_{p,w}$ be convex?

(b) Show that if X is a convex set, then $B_{p,w}$ is as well.

2.D.4^{A} Show that the budget set in Figure 2.D.4 is not convex.

2.E.1^{A} In text.

2.E.2^{B} In text.

2.E.3^{B} Use Propositions 2.E.1 to 2.E.3 to show that $p \cdot D_p x(p, w)\, p = -w$. Interpret.

2.E.4^{B} Show that if $x(p, w)$ is homogeneous of degree one with respect to w [i.e., $x(p, \alpha w) = \alpha x(p, w)$

for all $\alpha > 0$] and satisfies Walras' law, then $\varepsilon_{\ell w}(p, w) = 1$ for every ℓ. Interpret. Can you say something about $D_w x(p, w)$ and the form of the Engel functions and curves in this case?

2.E.5$^{\mathbf{B}}$ Suppose that $x(p, w)$ is a demand function which is homogeneous of degree one with respect to w and satisfies Walras' law and homogeneity of degree zero. Suppose also that all the cross-price effects are zero, that is $\partial x_\ell(p, w)/\partial p_k = 0$ whenever $k \neq \ell$. Show that this implies that for every ℓ, $x_\ell(p, w) = \alpha_\ell w/p_\ell$, where $\alpha_\ell > 0$ is a constant independent of (p, w).

2.E.6$^{\mathbf{A}}$ Verify that the conclusions of Propositions 2.E.1 to 2.E.3 hold for the demand function given in Exercise 2.E.1 when $\beta = 1$.

2.E.7$^{\mathbf{A}}$ A consumer in a two-good economy has a demand function $x(p, w)$ that satisfies Walras' law. His demand function for the first good is $x_1(p, w) = \alpha w/p_1$. Derive his demand function for the second good. Is his demand function homogeneous of degree zero?

2.E.8$^{\mathbf{B}}$ Show that the elasticity of demand for good ℓ with respect to price p_k, $\varepsilon_{\ell k}(p, w)$, can be written as $\varepsilon_{\ell k}(p, w) = d \ln (x_\ell(p, w))/d \ln (p_k)$, where $\ln(\cdot)$ is the natural logarithm function. Derive a similar expression for $\varepsilon_{\ell w}(p, w)$. Conclude that if we estimate the parameters $(\alpha_0, \alpha_1, \alpha_2, \gamma)$ of the equation $\ln (x_\ell(p, w)) = \alpha_0 + \alpha_1 \ln p_1 + \alpha_2 \ln p_2 + \gamma \ln w$, these parameter estimates provide us with estimates of the elasticities $\varepsilon_{\ell 1}(p, w)$, $\varepsilon_{\ell 2}(p, w)$, and $\varepsilon_{\ell w}(p, w)$.

2.F.1$^{\mathbf{B}}$ Show that for Walrasian demand functions, the definition of the weak axiom given in Definition 2.F.1 coincides with that in Definition 1.C.1.

2.F.2$^{\mathbf{B}}$ Verify the claim of Example 2.F.1.

2.F.3$^{\mathbf{B}}$ You are given the following partial information about a consumer's purchases. He consumes only two goods.

	Year 1		Year 2	
	Quantity	Price	Quantity	Price
Good 1	100	100	120	100
Good 2	100	100	?	80

Over what range of quantities of good 2 consumed in year 2 would you conclude:

(a) That his behaviour is inconsistent (i.e., in contradiction with the weak axiom)?

(b) That the consumer's consumption bundle in year 1 is revealed preferred to that in year 2?

(c) That the consumer's consumption bundle in year 2 is revealed preferred to that in year 1?

(d) That there is insufficient information to justify (a), (b), and/or (c)?

(e) That good 1 is an inferior good (at some price) for this consumer? Assume that the weak axiom is satisfied.

(f) That good 2 is an inferior good (at some price) for this consumer? Assume that the weak axiom is satisfied.

2.F.4$^{\mathbf{A}}$ Consider the consumption of a consumer in two different periods, period 0 and period 1. Period t prices, wealth, and consumption are p^t, w_t, and $x^t = x(p^t, w_t)$, respectively. It is often of applied interest to form an index measure of the quantity consumed by a consumer. The *Laspeyres* quantity index computes the change in quantity using period 0 prices as weights: $L_Q = (p^0 \cdot x^1)/(p^0 \cdot x^0)$. The *Paasche* quantity index instead uses period 1 prices as weights: $P_Q = (p^1 \cdot x^1)/(p^1 \cdot x^0)$. Finally, we could use the consumer's expenditure change: $E_Q = (p^1 \cdot x^1)/(p^0 \cdot x^0)$. Show the following:

(a) If $L_Q < 1$, then the consumer has a revealed preference for x^0 over x^1.

(b) If $P_Q > 1$, then the consumer has a revealed preference for x^1 over x^0.

(c) No revealed preference relationship is implied by either $E_Q > 1$ or $E_Q < 1$. Note that at the aggregate level, E_Q corresponds to the percentage change in gross national product.

2.F.5^C Suppose that $x(p, w)$ is a differentiable demand function that satisfies the weak axiom, Walras' law, and homogeneity of degree zero. Show that if $x(\cdot,\cdot)$ is homogeneous of degree one with respect to w [i.e., $x(p, \alpha w) = \alpha x(p, w)$ for all (p, w) and $\alpha > 0$], then the law of demand holds even for *uncompensated* price changes. If this is easier, establish only the infinitesimal version of this conclusion; that is, $dp \cdot D_p x(p, w)\, dp \leq 0$ for any dp.

2.F.6^A Suppose that $x(p, w)$ is homogeneous of degree zero. Show that the weak axiom holds if and only if for some $w > 0$ and all p, p' we have $p' \cdot x(p, w) > w$ whenever $p \cdot x(p', w) \leq w$ and $x(p', w) \neq x(p, w)$.

2.F.7^B In text.

2.F.8^A Let $\hat{s}_{\ell k}(p, w) = [p_k / x_\ell(p, w)] s_{\ell k}(p, w)$ be the substitution terms in elasticity form. Express $\hat{s}_{\ell k}(p, w)$ in terms of $\varepsilon_{\ell k}(p, w)$, $\varepsilon_{\ell w}(p, w)$, and $b_k(p, w)$.

2.F.9^B A symmetric $n \times n$ matrix A is negative definite if and only if $(-1)^k |A_{kk}| > 0$ for all $k \leq n$, where A_{kk} is the submatrix of A obtained by deleting the last $n - k$ rows and columns. For semidefiniteness of the symmetric matrix A, we replace the strict inequalities by weak inequalities and require that the weak inequalities hold for all matrices formed by permuting the rows and columns of A (see Section M.D of the Mathematical Appendix for details).

(a) Show that an arbitrary (possibly nonsymmetric) matrix A is negative definite (or semidefinite) if and only if $A + A^T$ is negative definite (or semidefinite). Show also that the above determinant condition (which can be shown to be necessary) is no longer sufficient in the nonsymmetric case.

(b) Show that for $L = 2$, the necessary and sufficient condition for the substitution matrix $S(p, w)$ of rank 1 to be negative semidefinite is that any diagonal entry (i.e., any own-price substitution effect) be negative.

2.F.10^B Consider the demand function in Exercise 2.E.1 with $\beta = 1$. Assume that $w = 1$.

(a) Compute the substitution matrix. Show that at $p = (1, 1, 1)$, it is negative semidefinite but not symmetric.

(b) Show that this demand function does not satisfy the weak axiom. [*Hint*: Consider the price vector $p = (1, 1, \varepsilon)$ and show that the substitution matrix is not negative semidefinite (for $\varepsilon > 0$ small).]

2.F.11^A Show that for $L = 2$, $S(p, w)$ is always symmetric. [*Hint*: Use Proposition 2.F.3.]

2.F.12^A Show that if the Walrasian demand function $x(p, w)$ is generated by a rational preference relation, than it must satisfy the weak axiom.

2.F.13^C Suppose that $x(p, w)$ may be multivalued.

(a) From the definition of the weak axiom given in Section 1.C, develop the generalization of Definition 2.F.1 for Walrasian demand correspondences.

(b) Show that if $x(p, w)$ satisfies this generalization of the weak axiom and Walras' law, then $x(\cdot)$ satisfies the following property:

(∗) For any $x \in x(p, w)$ and $x' \in x(p', w')$, if $p \cdot x' < w$, then $p \cdot x > w$.

(c) Show that the generalized weak axiom and Walras' law implies the following generalized version of the compensated law of demand: Starting from any initial position (p, w) with demand $x \in x(p, w)$, for any compensated price change to new prices p' and wealth level $w' = p' \cdot x$, we have

$$(p' - p) \cdot (x' - x) \leq 0$$

for all $x' \in x(p', w')$, with strict inequality if $x' \in x(p, w)$.

(d) Show that if $x(p, w)$ satisfies Walras' law and the generalized compensated law of demand defined in **(c)**, then $x(p, w)$ satisfies the generalized weak axiom.

2.F.14$^{\mathbf{A}}$ Show that if $x(p, w)$ is a Walrasian demand function that satisfies the weak axiom, then $x(p, w)$ must be homogeneous of degree zero.

2.F.15$^{\mathbf{B}}$ Consider a setting with $L = 3$ and a consumer whose consumption set is $\mathbb{R}^3$. The consumer's demand function $x(p, w)$ satisfies homogeneity of degree zero, Walras' law and (fixing $p_3 = 1$) has

$$x_1(p, w) = -p_1 + p_2$$

and

$$x_2(p, w) = -p_2.$$

Show that this demand function satisfies the weak axiom by demonstrating that its substitution matrix satisfies $v \cdot S(p, w)\, v < 0$ for all $v \neq \alpha p$. [*Hint*: Use the matrix results recorded in Section M.D of the Mathematical Appendix.] Observe then that the substitution matrix is not symmetric. (*Note*: The fact that we allow for negative consumption levels here is not essential for finding a demand function that satisfies the weak axiom but whose substitution matrix is not symmetric; with a consumption set allowing only for nonnegative consumption levels, however, we would need to specify a more complicated demand function.)

2.F.16$^{\mathbf{B}}$ Consider a setting where $L = 3$ and a consumer whose consumption set is $\mathbb{R}^3$. Suppose that his demand function $x(p, w)$ is

$$x_1(p, w) = \frac{p_2}{p_3},$$

$$x_2(p, w) = -\frac{p_1}{p_3},$$

$$x_3(p, w) = \frac{w}{p_3}.$$

(a) Show that $x(p, w)$ is homogeneous of degree zero in (p, w) and satisfies Walras' law.

(b) Show that $x(p, w)$ violates the weak axiom.

(c) Show that $v \cdot S(p, w)\, v = 0$ for all $v \in \mathbb{R}^3$.

2.F.17$^{\mathbf{B}}$ In an L-commodity world, a consumer's Walrasian demand function is

$$x_k(p, w) = \frac{w}{\left(\sum_{\ell=1}^{L} p_\ell\right)} \quad \text{for } k = 1, \ldots, L.$$

(a) Is this demand function homogeneous of degree zero in (p, w)?

(b) Does it satisfy Walras' law?

(c) Does it satisfy the weak axiom?

(d) Compute the Slutsky substitution matrix for this demand function. Is it negative semidefinite? Negative definite? Symmetric?

CHAPTER

3

Classical Demand Theory

3.A Introduction

In this chapter, we study the classical, preference-based approach to consumer demand.

We begin in Section 3.B by introducing the consumer's preference relation and some of its basic properties. We assume throughout that this preference relation is *rational*, offering a complete and transitive ranking of the consumer's possible consumption choices. We also discuss two properties, *monotonicity* (or its weaker version, *local nonsatiation*) and *convexity*, that are used extensively in the analysis that follows.

Section 3.C considers a technical issue: the existence and continuity properties of utility functions that represent the consumer's preferences. We show that not all preference relations are representable by a utility function, and we then formulate an assumption on preferences, known as *continuity*, that is sufficient to guarantee the existence of a (continuous) utility function.

In Section 3.D, we begin our study of the consumer's decision problem by assuming that there are L commodities whose prices she takes as fixed and independent of her actions (the *price-taking assumption*). The consumer's problem is framed as one of *utility maximization* subject to the constraints embodied in the Walrasian budget set. We focus our study on two objects of central interest: the consumer's optimal choice, embodied in the *Walrasian* (or *market* or *ordinary*) *demand correspondence*, and the consumer's optimal utility value, captured by the *indirect utility function*.

Section 3.E introduces the consumer's *expenditure minimization problem*, which bears a close relation to the consumer's goal of utility maximization. In parallel to our study of the demand correspondence and value function of the utility maximization problem, we study the equivalent objects for expenditure minimization. They are known, respectively, as the *Hicksian* (or *compensated*) *demand correspondence* and the *expenditure function*. We also provide an initial formal examination of the relationship between the expenditure minimization and utility maximization problems.

In Section 3.F, we pause for an introduction to the mathematical underpinnings of duality theory. This material offers important insights into the structure of

preference-based demand theory. Section 3.F may be skipped without loss of continuity in a first reading of the chapter. Nevertheless, we recommend the study of its material.

Section 3.G continues our analysis of the utility maximization and expenditure minimization problems by establishing some of the most important results of demand theory. These results develop the fundamental connections between the demand and value functions of the two problems.

In Section 3.H, we complete the study of the implications of the preference-based theory of consumer demand by asking how and when we can recover the consumer's underlying preferences from her demand behavior, an issue traditionally known as the *integrability problem.* In addition to their other uses, the results presented in this section tell us that the properties of consumer demand identified in Sections 3.D to 3.G as *necessary* implications of preference-maximizing behavior are also *sufficient* in the sense that any demand behavior satisfying these properties can be rationalized as preference-maximizing behavior.

The results in Sections 3.D to 3.H also allow us to compare the implications of the preference-based approach to consumer demand with the choice-based theory studied in Section 2.F. Although the differences turn out to be slight, the two approaches are not equivalent; the choice-based demand theory founded on the weak axiom of revealed preference imposes fewer restrictions on demand than does the preference-based theory studied in this chapter. The extra condition added by the assumption of rational preferences turns out to be the *symmetry* of the Slutsky matrix. As a result, we conclude that satisfaction of the weak axiom does not ensure the existence of a rationalizing preference relation for consumer demand.

Although our analysis in Sections 3.B to 3.H focuses entirely on the positive (i.e., descriptive) implications of the preference-based approach, one of the most important benefits of the latter is that it provides a framework for normative, or *welfare*, analysis. In Section 3.I, we take a first look at this subject by studying the effects of a price change on the consumer's welfare. In this connection, we discuss the use of the traditional concept of Marshallian surplus as a measure of consumer welfare.

We conclude in Section 3.J by returning to the choice-based approach to consumer demand. We ask whether there is some strengthening of the weak axiom that leads to a choice-based theory of consumer demand equivalent to the preference-based approach. As an answer, we introduce the *strong axiom of revealed preference* and show that it leads to demand behavior that is consistent with the existence of underlying preferences.

Appendix A discusses some technical issues related to the continuity and differentiability of Walrasian demand.

For further reading, see the thorough treatment of classical demand theory offered by Deaton and Muellbauer (1980).

3.B Preference Relations: Basic Properties

In the classical approach to consumer demand, the analysis of consumer behavior begins by specifying the consumer's preferences over the commodity bundles in the consumption set $X \subset \mathbb{R}^L_+$.

The consumer's preferences are captured by a preference relation $\succsim$ (an "at-least-as-good-as" relation) defined on X that we take to be *rational* in the sense introduced in Section 1.B; that is, $\succsim$ is *complete* and *transitive*. For convenience, we repeat the formal statement of this assumption from Definition 1.B.1.[1]

Definition 3.B.1: The preference relation $\succsim$ on X is *rational* if it possesses the following two properties:

(i) *Completeness*. For all $x, y \in X$, we have $x \succsim y$ or $y \succsim x$ (or both).
(ii) *Transitivity*. For all $x, y, z \in X$, if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.

In the discussion that follows, we also use two other types of assumptions about preferences: *desirability* assumptions and *convexity* assumptions.

(i) *Desirability assumptions.* It is often reasonable to assume that larger amounts of commodities are preferred to smaller ones. This feature of preferences is captured in the assumption of monotonicity. For Definition 3.B.2, we assume that the consumption of larger amounts of goods is always feasible in principle; that is, if $x \in X$ and $y \geq x$, then $y \in X$.

Definition 3.B.2: The preference relation $\succsim$ on X is *monotone* if $x \in X$ and $y \gg x$ implies $y \succ x$. It is *strongly monotone* if $y \geq x$ and $y \neq x$ imply that $y \succ x$.

The assumption that preferences are monotone is satisfied as long as commodities are "goods" rather than "bads". Even if some commodity is a bad, however, we may still be able to view preferences as monotone because it is often possible to redefine a consumption activity in a way that satisfies the assumption. For example, if one commodity is garbage, we can instead define the individual's consumption over the "absence of garbage".[2]

Note that if $\succsim$ is monotone, we may have indifference with respect to an increase in the amount of some but not all commodities. In contrast, strong montonicity says that if y is larger than x for *some* commodity and is no less for any other, then y is strictly preferred to x.

For much of the theory, however, a weaker desirability assumption than monotonicity, known as *local nonsatiation*, actually suffices.

Definition 3.B.3: The preference relation $\succsim$ on X is *locally nonsatiated* if for every $x \in X$ and every $\varepsilon > 0$, there is $y \in X$ such that $\|y - x\| \leq \varepsilon$ and $y \succ x$.[3]

The test for locally nonsatiated preferences is depicted in Figure 3.B.1 for the case in which $X = \mathbb{R}^L_+$. It says that for any consumption bundle $x \in \mathbb{R}^L_+$ and any arbitrarily

---

1. See Section 1.B for a thorough discussion of these properties.

2. It is also sometimes convenient to view preferences as defined over the level of goods available for consumption (the stocks of goods on hand), rather than over the consumption levels themselves. In this case, if the consumer can freely dispose of any unwanted commodities, her preferences over the level of commodities on hand are monotone as long as some good is always desirable.

3. $\|x - y\|$ is the Euclidean distance between points x and y; that is, $\|x - y\| = [\sum_{\ell=1}^{L} (x_\ell - y_\ell)^2]^{1/2}$.

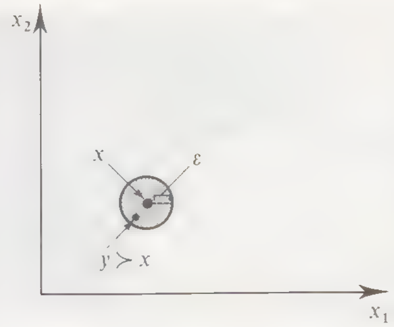


Figure 3.B.1
The test for local nonsatiation.

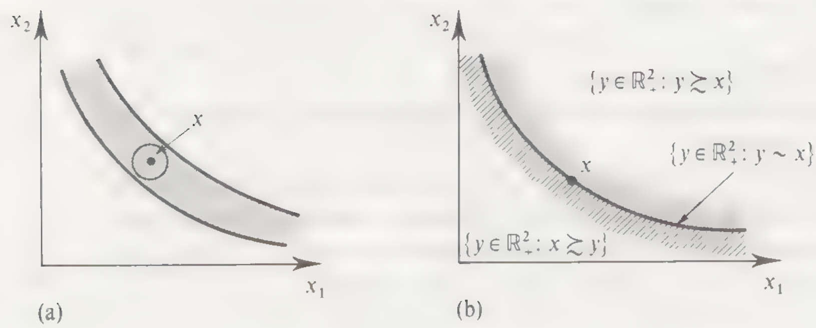


Figure 3.B.2
(a) A thick indifference set violates local nonsatiation.
(b) Preferences compatible with local nonsatiation.

small distance away from x, denoted by $\varepsilon > 0$, there is another bundle $y \in \mathbb{R}^L_+$ within this distance from x that is preferred to x. Note that the bundle y may even have less of every commodity than x, as shown in the figure. Nonetheless, when $X = \mathbb{R}^L_+$ local nonsatiation rules out the extreme situation in which all commodities are bads, since in that case no consumption at all (the point $x = 0$) would be a satiation point.

Exercise 3.B.1: Show the following:

(a) If $\succsim$ is strongly monotone, then it is monotone.
(b) If $\succsim$ is monotone, then it is locally nonsatiated.

Given the preference relation $\succsim$ and a consumption bundle x, we can define three related sets of consumption bundles. The *indifference set* containing point x is the set of all bundles that are indifferent to x; formally, it is $\{y \in X : y \sim x\}$. The *upper contour set* of bundle x is the set of all bundles that are at least as good as x: $\{y \in X : y \succsim x\}$. The *lower contour set* of x is the set of all bundles that x is at least as good as: $\{y \in X : x \succsim y\}$.

One implication of local nonsatiation (and, hence, of monotonicity) is that it rules out "thick" indifference sets. The indifference set in Figure 3.B.2(a) cannot satisfy local nonsatiation because, if it did, there would be a better point than x within the circle drawn. In contrast, the indifference set in Figure 3.B.2(b) is compatible with local nonsatiation. Figure 3.B.2(b) also depicts the upper and lower contour sets of x.

(ii) *Convexity assumptions.* A second significant assumption, that of convexity of $\succsim$, concerns the trade-offs that the consumer is willing to make among different goods.

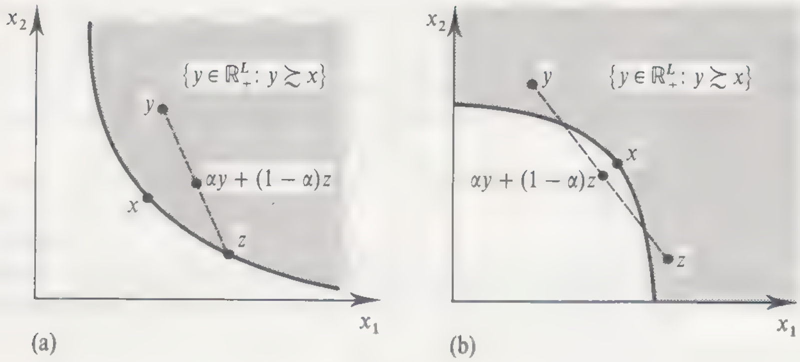


Figure 3.B.3
(a) Convex preferences.
(b) Nonconvex preferences.

Definition 3.B.4: The preference relation $\succsim$ on X is *convex* if for every $x \in X$, the upper contour set $\{y \in X: y \succsim x\}$ is convex; that is, if $y \succsim x$ and $z \succsim x$, then $\alpha y + (1-\alpha)z \succsim x$ for any $\alpha \in [0, 1]$.

Figure 3.B.3(a) depicts a convex upper contour set; Figure 3.B.3(b) shows an upper contour set that is not convex.

Convexity is a strong but central hypothesis in economics. It can be interpreted in terms of *diminishing marginal rates of substitution*: That is, with convex preferences, from any initial consumption situation x, and for any two commodities, it takes increasingly larger amounts of one commodity to compensate for successive unit losses of the other.[4]

Convexity can also be viewed as the formal expression of a basic inclination of economic agents for diversification. Indeed, under convexity, if x is indifferent to y, then $\frac{1}{2}x + \frac{1}{2}y$, the half–half mixture of x and y, cannot be worse than either x or y. In Chapter 6, we shall give a diversification interpretation in terms of behavior under uncertainty. A taste for diversification is a realistic trait of economic life. Economic theory would be in serious difficulty if this postulated propensity for diversification did not have significant descriptive content. But there is no doubt that one can easily think of choice situations where it is violated. For example, you may like both milk and orange juice but get less pleasure from a mixture of the two.

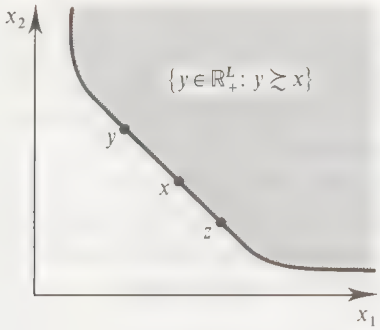
Definition 3.B.4 has been stated for a general consumption set X. But de facto, the convexity assumption can hold only if X is convex. Thus, the hypothesis rules out commodities being consumable only in integer amounts or situations such as that presented in Figure 2.C.3.

Although the convexity assumption on preferences may seem strong, this appearance should be qualified in two respects: First, a good number (although not all) of the results of this chapter extend without modification to the nonconvex case. Second, as we show in Appendix A of Chapter 4 and in Section 17.I, nonconvexities can often be incorporated into the theory by exploiting regularizing aggregation effects across consumers.

We also make use at times of a strengthening of the convexity assumption.

Definition 3.B.5: The preference relation $\succsim$ on X is *strictly convex* if for every x, we have that $y \succsim x$, $z \succsim x$, and $y \neq z$ implies $\alpha y + (1-\alpha)z \succ x$ for all $\alpha \in (0, 1)$.

4. More generally, convexity is equivalent to a diminishing marginal rate of substitution between any two goods, provided that we allow for "composite commodities" formed from linear combinations of the L basic commodities.



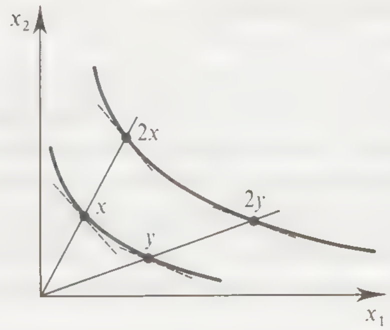


Figure 3.B.4 (left)

A convex, but not strictly convex, preference relation.

Figure 3.B.5 (right)

Homothetic preferences.

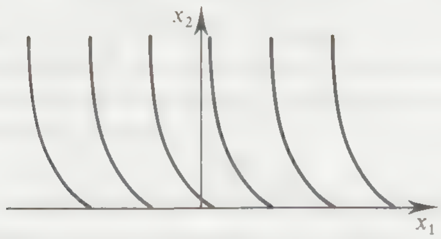


Figure 3.B.6

Quasilinear preferences.

Figure 3.B.3(a) showed strictly convex preferences. In Figure 3.B.4, on the other hand, the preferences, although convex, are not strictly convex.

In applications (particularly those of an econometric nature), it is common to focus on preferences for which it is possible to deduce the consumer's entire preference relation from a single indifference set. Two examples are the classes of *homothetic* and *quasilinear* preferences.

Definition 3.B.6: A monotone preference relation $\succsim$ on $X = \mathbb{R}^L_+$ is *homothetic* if all indifference sets are related by proportional expansion along rays; that is, if $x \sim y$, then $\alpha x \sim \alpha y$ for any $\alpha \geq 0$.

Figure 3.B.5 depicts a homothetic preference relation.

Definition 3.B.7: The preference relation $\succsim$ on $X = (-\infty, \infty) \times \mathbb{R}^{L-1}_+$ is *quasilinear* with respect to commodity 1 (called, in this case, the *numeraire* commodity) if[5]

(i) All the indifference sets are parallel displacements of each other along the axis of commodity 1. That is, if $x \sim y$, then $(x + \alpha e_1) \sim (y + \alpha e_1)$ for $e_1 = (1, 0, \ldots, 0)$ and any $\alpha \in \mathbb{R}$.

(ii) Good 1 is desirable; that is, $x + \alpha e_1 \succ x$ for all x and $\alpha > 0$.

Note that, in Definition 3.B.7, we assume that there is no lower bound on the possible consumption of the first commodity [the consumption set is $(-\infty, \infty) \times \mathbb{R}^{L-1}_+$]. This assumption is convenient in the case of quasilinear preferences (Exercise 3.D.4 will illustate why). Figure 3.B.6 shows a quasilinear preference relation.

5. More generally, preferences can be quasilinear with respect to any commodity ℓ.

3.C Preference and Utility

For analytical purposes, it is very helpful if we can summarize the consumer's preferences by means of a utility function because mathematical programming techniques can then be used to solve the consumer's problem. In this section, we study when this can be done. Unfortunately, with the assumptions made so far, a rational preference relation need not be representable by a utility function. We begin with an example illustrating this fact and then introduce a weak, economically natural assumption (called *continuity*) that guarantees the existence of a utility representation.

Example 3.C.1: *The Lexicographic Preference Relation.* For simplicity, assume that $X = \mathbb{R}^2_+$. Define $x \succsim y$ if either "$x_1 > y_1$" or "$x_1 = y_1$ and $x_2 \geq y_2$." This is known as the *lexicographic preference relation.* The name derives from the way a dictionary is organized; that is, commodity 1 has the highest priority in determining the preference ordering, just as the first letter of a word does in the ordering of a dictionary. When the level of the first commodity in two commodity bundles is the same, the amount of the second commodity in the two bundles determines the consumer's preferences. In Exercise 3.C.1, you are asked to verify that the lexicographic ordering is complete, transitive, strongly monotone, and strictly convex. Nevertheless, it can be shown that no utility function exists that represents this preference ordering. This is intuitive. With this preference ordering, no two distinct bundles are indifferent; indifference sets are singletons. Therefore, we have two dimensions of distinct indifference sets. Yet, each of these indifference sets must be assigned, in an order-preserving way, a different utility number from the one-dimensional real line. In fact, a somewhat subtle argument is actually required to establish this claim rigorously. It is given, for the more advanced reader, in the following paragraph.

Suppose there is a utility function $u(\cdot)$. For every x_1, we can pick a rational number $r(x_1)$ such that $u(x_1, 2) > r(x_1) > u(x_1, 1)$. Note that because of the lexicographic character of preferences, $x_1 > x'_1$ implies $r(x_1) > r(x'_1)$ [since $r(x_1) > u(x_1, 1) > u(x'_1, 2) > r(x'_1)$]. Therefore, $r(\cdot)$ provides a one-to-one function from the set of real numbers (which is uncountable) to the set of rational numbers (which is countable). This is a mathematical impossibility. Therefore, we conclude that there can be no utility function representing these preferences.

■

The assumption that is needed to ensure the existence of a utility function is that the preference relation be continuous.

Definition 3.C.1: The preference relation $\succsim$ on X is *continuous* if it is preserved under limits. That is, for any sequence of pairs $\{(x^n, y^n)\}_{n=1}^{\infty}$ with $x^n \succsim y^n$ for all n, $x = \lim_{n\to\infty} x^n$, and $y = \lim_{n\to\infty} y^n$, we have $x \succsim y$.

Continuity says that the consumer's preferences cannot exhibit "jumps," with, for example, the consumer preferring each element in sequence $\{x^n\}$ to the corresponding element in sequence $\{y^n\}$ but suddenly reversing her preference at the limiting points of these sequences x and y.

An equivalent way to state this notion of continuity is to say that for all x, the upper contour set $\{y \in X: y \succsim x\}$ and the lower contour set $\{y \in X: x \succsim y\}$ are both *closed*; that is, they include their boundaries. Definition 3.C.1 implies that for any sequence of points $\{y^n\}_{n=1}^{\infty}$ with $x \succsim y^n$ for all n and $y = \lim_{n\to\infty} y^n$, we have $x \succsim y$ (just let $x^n = x$ for all n). Hence, continuity as defined in Definition 3.C.1 implies that the lower contour set is closed; the same is implied for the upper contour set. The reverse argument, that closedness of the lower and upper contour sets implies that Definition 3.C.1 holds, is more advanced and is left as an exercise (Exercise 3.C.3).

Example 3.C.1 continued: Lexicographic preferences are not continuous. To see this, consider the sequence of bundles $x^n = (1/n, 0)$ and $y^n = (0, 1)$. For every n, we have $x^n \succ y^n$. But $\lim_{n\to\infty} y^n = (0, 1) \succ (0, 0) = \lim_{n\to\infty} x^n$. In words, as long as the first component of x is larger than that of y, x is preferred to y even if y_2 is much larger than x_2. But as soon as the first components become equal, only the second components are relevant, and so the preference ranking is reversed at the limit points of the sequence. ■

It turns out that the continuity of $\succsim$ is sufficient for the existence of a utility function representation. In fact, it guarantees the existence of a *continuous* utility function.

Proposition 3.C.1: Suppose that the rational preference relation $\succsim$ on X is continuous. Then there is a continuous utility function $u(x)$ that represents $\succsim$.

Proof: For the case of $X = \mathbb{R}^L_+$ and a monotone preference relation, there is a relatively simple and intuitive proof that we present here with the help of Figure 3.C.1.

Denote the diagonal ray in $\mathbb{R}^L_+$ (the locus of vectors with all L components equal) by Z. It will be convenient to let e designate the L-vector whose elements are all equal to 1. Then $\alpha e \in Z$ for all nonnegative scalars $\alpha \geq 0$.

Note that for every $x \in \mathbb{R}^L_+$, monotonicity implies that $x \succsim 0$. Also note that for any $\bar{\alpha}$ such that $\bar{\alpha}e \gg x$ (as drawn in the figure), we have $\bar{\alpha}e \succsim x$. Monotonicity and continuity can then be shown to imply that there is a unique value $\alpha(x) \in [0, \bar{\alpha}]$ such that $\alpha(x)e \sim x$.

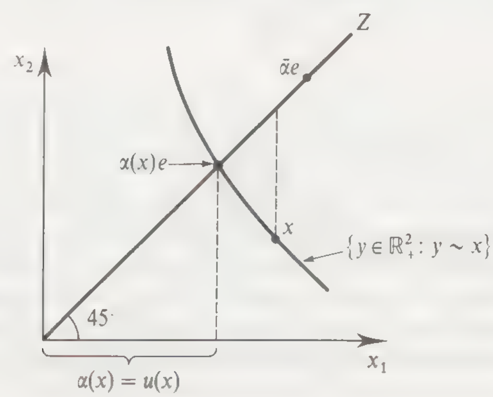


Figure 3.C.1 Construction of a utility function.

Formally, this can be shown as follows: By continuity, the upper and lower contour sets of x are closed. Hence, the sets $A^+ = \{\alpha \in \mathbb{R}_+ : \alpha e \succsim x\}$ and $A^- = \{\alpha \in \mathbb{R}_+ : x \succsim \alpha e\}$ are nonempty and closed. Note that by completeness of $\succsim$, $\mathbb{R}_+ \subset (A^+ \cup A^-)$. The nonemptiness and closedness of A^+ and A^-, along with the fact that $\mathbb{R}_+$ is connected, imply that $A^+ \cap A^- \neq \varnothing$. Thus, there exists a scalar α such that $\alpha e \sim x$. Furthermore, by monotonicity, $\alpha_1 e \succ \alpha_2 e$ whenever $\alpha_1 > \alpha_2$. Hence, there can be at most one scalar satisfying $\alpha e \sim x$. This scalar is $\alpha(x)$.

We now take $\alpha(x)$ as our utility function; that is, we assign a utility value $u(x) = \alpha(x)$ to every x. This utility level is also depicted in Figure 3.C.1. We need to check two properties of this function: that it represents the preference $\succsim$ [i.e., that $\alpha(x) \geq \alpha(y) \Leftrightarrow x \succsim y$] and that it is a continuous function. The latter argument is more advanced, and therefore we present it in small type.

That $\alpha(x)$ represents preferences follows from its construction. Formally, suppose first that $\alpha(x) \geq \alpha(y)$. By monotonicity, this implies that $\alpha(x)e \succsim \alpha(y)e$. Since $x \sim \alpha(x)e$ and $y \sim \alpha(y)e$, we have $x \succsim y$. Suppose, on the other hand, that $x \succsim y$. Then $\alpha(x)e \sim x \succsim y \sim \alpha(y)e$; and so by monotonicity, we must have $\alpha(x) \geq \alpha(y)$. Hence, $\alpha(x) \geq \alpha(y) \Leftrightarrow x \succsim y$.

We now argue that $\alpha(x)$ is a continuous function at all x; that is, for any sequence $\{x^n\}_{n=1}^{\infty}$ with $x = \lim_{n\to\infty} x^n$, we have $\lim_{n\to\infty} \alpha(x^n) = \alpha(x)$. Hence, consider a sequence $\{x^n\}_{n=1}^{\infty}$ such that $x = \lim_{n\to\infty} x^n$.

We note first that the sequence $\{\alpha(x^n)\}_{n=1}^{\infty}$ must have a convergent subsequence. By monotonicity, for any $\varepsilon > 0$, $\alpha(x')$ lies in a compact subset of $\mathbb{R}_+$, $[\alpha_0, \alpha_1]$, for all x' such that $\|x' - x\| \leq \varepsilon$ (see Figure 3.C.2). Since $\{x^n\}_{n=1}^{\infty}$ converges to x, there exists an N such that $\alpha(x^n)$

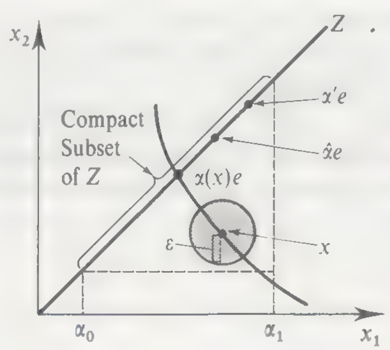


Figure 3.C.2
Proof that the constructed utility function is continuous.

lies in this compact set for all $n > N$. But any infinite sequence that lies in a compact set must have a convergent subsequence (see Section M.F of the Mathematical Appendix).

What remains is to establish that all convergent subsequences of $\{\alpha(x^n)\}_{n=1}^{\infty}$ converge to $\alpha(x)$. To see this, suppose otherwise: that there is some strictly increasing function $m(\cdot)$ that assigns to each positive integer n a positive integer $m(n)$ and for which the subsequence $\{\alpha(x^{m(n)})\}_{n=1}^{\infty}$ converges to $\alpha' \neq \alpha(x)$. We first show that $\alpha' > \alpha(x)$ leads to a contradiction. To begin, note that monotonicity would then imply that $\alpha' e \succ \alpha(x)e$. Now, let $\hat{\alpha} = \frac{1}{2}[\alpha' + \alpha(x)]$. The point $\hat{\alpha}e$ is the midpoint on Z between $\alpha' e$ and $\alpha(x)e$ (see Figure 3.C.2). By monotonicity, $\hat{\alpha}e \succ \alpha(x)e$. Now, since $\alpha(x^{m(n)}) \to \alpha' > \hat{\alpha}$, there exists an $\bar{N}$ such that for all $n > \bar{N}$, $\alpha(x^{m(n)}) > \hat{\alpha}$.

Hence, for all such n, $x^{m(n)} \sim \alpha(x^{m(n)})e \succ \hat{\alpha}e$ (where the latter relation follows from monotonicity). Because preferences are continuous, this would imply that $x \succsim \hat{\alpha}e$. But since $x \sim \alpha(x)e$, we get $\alpha(x)e \succsim \hat{\alpha}e$, which is a contradiction. The argument ruling out $\alpha' < \alpha(x)$ is similar. Thus, since all convergent subsequences of $\{\alpha(x^n)\}_{n=1}^{\infty}$ must converge to $\alpha(x)$, we have $\lim_{n \to \infty} \alpha(x^n) = \alpha(x)$, and we are done.

■

From now on, we assume that the consumer's preference relation is continuous and hence representable by a continuous utility function. As we noted in Section 1.B, the utility function $u(\cdot)$ that represents a preference relation $\succsim$ is not unique; any strictly increasing transformation of $u(\cdot)$, say $v(x) = f(u(x))$, where $f(\cdot)$ is a strictly increasing function, also represents $\succsim$. Proposition 3.C.1 tells us that if $\succsim$ is continuous, there exists *some* continuous utility function representing $\succsim$. But not all utility functions representing $\succsim$ are continuous; any strictly increasing but discontinuous transformation of a continuous utility function also represents $\succsim$.

For analytical purposes, it is also convenient if $u(\cdot)$ can be assumed to be differentiable. It is possible, however, for continuous preferences *not* to be representable by a differentiable utility function. The simplest example, shown in Figure 3.C.3, is the case of *Leontief* preferences, where $x'' \succsim x'$ if and only if $\text{Min}\,\{x_1'', x_2''\} \geq \text{Min}\,\{x_1', x_2'\}$. The nondifferentiability arises because of the kink in indifference curves when $x_1 = x_2$.

innymvd

Whenever convenient in the discussion that follows, we nevertheless assume utility functions to be twice continuously differentiable. It is possible to give a condition purely in terms of preferences that implies this property, but we shall not do so here. Intuitively, what is required is that indifference sets be smooth surfaces that fit together nicely so that the rates at which commodities substitute for each other depend differentiably on the consumption levels.

Restrictions on preferences translate into restrictions on the form of utility functions. The property of monotonicity, for example, implies that the utility function is increasing: $u(x) > u(y)$ if $x \gg y$.

The property of convexity of preferences, on the other hand, implies that $u(\cdot)$ is *quasiconcave* [and, similarly, strict convexity of preferences implies strict quasiconcavity of $u(\cdot)$]. The utility function $u(\cdot)$ is quasiconcave if the set $\{y \in \mathbb{R}_+^L : u(y) \geq u(x)\}$ is convex for all x or, equivalently, if $u(\alpha x + (1 - \alpha)y) \geq \text{Min}\,\{u(x), u(y)\}$ for

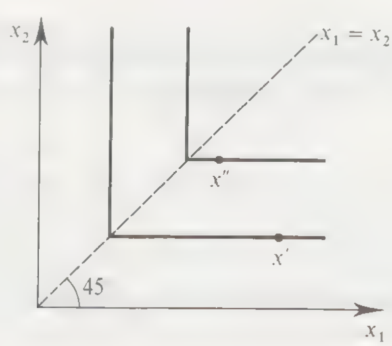


Figure 3.C.3
Leontief preferences cannot be represented by a differentiable utility function.

any x, y and all $\alpha \in [0, 1]$. [If the inequality is strict for all $x \neq y$ and $\alpha \in (0, 1)$ then $u(\cdot)$ is strictly quasiconcave; for more on quasiconcavity and strict quasiconcavity see Section M.C of the Mathematical Appendix.] Note, however, that convexity of $\succsim$ does *not* imply the stronger property that $u(\cdot)$ is concave [that $u(\alpha x + (1 - \alpha)y) \geq \alpha u(x) + (1 - \alpha)u(y)$ for any x, y and all $\alpha \in [0, 1]$]. In fact, although this is a somewhat fine point, there may not be *any* concave utility function representing a particular convex preference relation $\succsim$.

In Exercise 3.C.5, you are asked to prove two other results relating utility representations and underlying preference relations:

(i) A continuous $\succsim$ on $X = \mathbb{R}^L_+$ is homothetic if and only if it admits a utility function $u(x)$ that is homogeneous of degree one [i.e., such that $u(\alpha x) = \alpha u(x)$ for all $\alpha > 0$].

(ii) A continuous $\succsim$ on $(-\infty, \infty) \times \mathbb{R}^{L-1}_+$ is quasilinear with respect to the first commodity if and only if it admits a utility function $u(x)$ of the form $u(x) = x_1 + \phi(x_2, \ldots, x_L)$.

It is important to realize that although monotonicity and convexity of $\succsim$ imply that *all* utility functions representing $\succsim$ are increasing and quasiconcave, (i) and (ii) merely say that there is at *least one* utility function that has the specified form. Increasingness and quasiconcavity are *ordinal* properties of $u(\cdot)$; they are preserved for any arbitrary increasing transformation of the utility index. In contrast, the special forms of the utility representations in (i) and (ii) are not preserved; they are *cardinal* properties that are simply convenient choices for a utility representation.[6]

3.D The Utility Maximization Problem

We now turn to the study of the consumer's decision problem. We assume throughout that the consumer has a rational, continuous, and locally nonsatiated preference relation, and we take $u(x)$ to be a continuous utility function representing these preferences. For the sake of concreteness, we also assume throughout the remainder of the chapter that the consumption set is $X = \mathbb{R}^L_+$.

The consumer's problem of choosing her most preferred consumption bundle given prices $p \gg 0$ and wealth level $w > 0$ can now be stated as the following *utility maximization problem* (*UMP*):

$$\begin{aligned} \underset{x \geq 0}{\text{Max}} \quad & u(x) \\ \text{s.t. } & p \cdot x \leq w. \end{aligned}$$

In the UMP, the consumer chooses a consumption bundle in the Walrasian budget set $B_{p,w} = \{x \in \mathbb{R}^L_+ : p \cdot x \leq w\}$ to maximize her utility level. We begin with the results stated in Proposition 3.D.1.

Proposition 3.D.1: If $p \gg 0$ and $u(\cdot)$ is continuous, then the utility maximization problem has a solution.

6. Thus, in this sense, continuity is also a cardinal property of utility functions. See also the discussion of ordinal and cardinal properties of utility representations in Section 1.B.

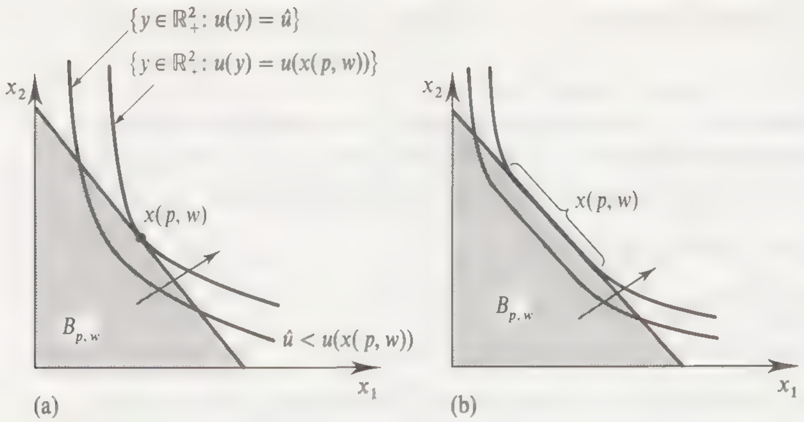


Figure 3.D.1
The utility maximization problem (UMP).
(a) Single solution.
(b) Multiple solutions.

Proof: If $p \gg 0$, then the budget set $B_{p,w} = \{x \in \mathbb{R}^L_+: p \cdot x \le w\}$ is a compact set because it is both bounded [for any $\ell = 1,\dots,L$, we have $x_\ell \le (w/p_\ell)$ for all $x \in B_{p,w}$] and closed. The result follows from the fact that a continuous function always has a maximum value on any compact set (set Section M.F. of the Mathematical Appendix). ■

With this result, we now focus our attention on the properties of two objects that emerge from the UMP: the consumer's set of optimal consumption bundles (the solution set of the UMP) and the consumer's maximal utility value (the value function of the UMP).

The Walrasian Demand Correspondence/Function

The rule that assigns the set of optimal consumption vectors in the UMP to each price–wealth situation $(p, w) \gg 0$ is denoted by $x(p, w) \in \mathbb{R}^L_+$ and is known as the *Walrasian* (or *ordinary* or *market*) *demand correspondence*. An example for $L = 2$ is depicted in Figure 3.D.1(a), where the point $x(p, w)$ lies in the indifference set with the highest utility level of any point in $B_{p,w}$. Note that, as a general matter, for a given $(p, w) \gg 0$ the optimal set $x(p, w)$ may have more than one element, as shown in Figure 3.D.1(b). When $x(p, w)$ is single-valued for all (p, w), we refer to it as the *Walrasian* (or *ordinary* or *market*) *demand function.*[7]

The properties of $x(p, w)$ stated in Proposition 3.D.2 follow from direct examination of the UMP.

Proposition 3.D.2: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. Then the Walrasian demand correspondence $x(p, w)$ possesses the following properties:

7. This demand function has also been called the *Marshallian demand function.* However, this terminology can create confusion, and so we do not use it here. In Marshallian partial equilibrium analysis (where wealth effects are absent), all the different kinds of demand functions studied in this chapter coincide, and so it is not clear which of these demand functions would deserve the Marshall name in the more general setting.

(i) *Homogeneity of degree zero in* (p, w)*:* $x(\alpha p, \alpha w) = x(p, w)$ for any p, w and scalar $\alpha > 0$.

(ii) *Walras' law:* $p \cdot x = w$ for all $x \in x(p, w)$.

(iii) *Convexity/uniqueness:* If $\succsim$ is convex, so that $u(\cdot)$ is quasiconcave, then $x(p, w)$ is a convex set. Moreover, if $\succsim$ is *strictly convex*, so that $u(\cdot)$ is strictly quasiconcave, then $x(p, w)$ consists of a single element.

Proof: We establish each of these properties in turn.

(i) For homogeneity, note that for any scalar $\alpha > 0$,

$$\{x \in \mathbb{R}^L_+ : \alpha p \cdot x \leq \alpha w\} = \{x \in \mathbb{R}^L_+ : p \cdot x \leq w\};$$

that is, the set of feasible consumption bundles in the UMP does not change when all prices and wealth are multiplied by a constant $\alpha > 0$. The set of utility-maximizing consumption bundles must therefore be the same in these two circumstances, and so $x(p, w) = x(\alpha p, \alpha w)$. Note that this property does not require any assumptions on $u(\cdot)$.

(ii) Walras' law follows from local nonsatiation. If $p \cdot x < w$ for some $x \in x(p, w)$, then there must exist another consumption bundle y sufficiently close to x with both $p \cdot y < w$ and $y \succ x$ (see Figure 3.D.2). But this would contradict x being optimal in the UMP.

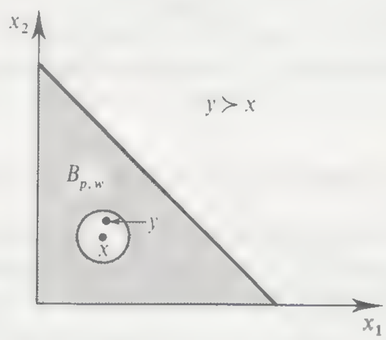


Figure 3.D.2
Local nonsatiation implies Walras' law.

(iii) Suppose that $u(\cdot)$ is quasiconcave and that there are two bundles x and x', with $x \neq x'$, both of which are elements of $x(p, w)$. To establish the result, we show that $x'' = \alpha x + (1 - \alpha)x'$ is an element of $x(p, w)$ for any $\alpha \in [0,1]$. To start, we know that $u(x) = u(x')$. Denote this utility level by u^*. By quasiconcavity, $u(x'') \geq u^*$ [see Figure 3.D.3(a)]. In addition, since $p \cdot x \leq w$ and $p \cdot x' \leq w$, we also have

$$p \cdot x'' = p \cdot [\alpha x + (1 - \alpha)x'] \leq w.$$

Therefore, x'' is a feasible choice in the UMP (put simply, x'' is feasible because $B_{p,w}$ is a convex set). Thus, since $u(x'') \geq u^*$ and x'' is feasible, we have $x'' \in x(p, w)$. This establishes that $x(p, w)$ is a convex set if $u(\cdot)$ is quasiconcave.

Suppose now that $u(\cdot)$ is *strictly* quasiconcave. Following the same argument but using strict quasiconcavity, we can establish that x'' is a feasible choice and that $u(x'') > u^*$ for all $\alpha \in (0,1)$. Because this contradicts the assumption that x and x' are elements of $x(p, w)$, we conclude that there can be at most one element in $x(p, w)$. Figure 3.D.3(b) illustrates this argument. Note the difference from Figure 3.D.3(a) arising from the strict quasiconcavity of $u(x)$. ■

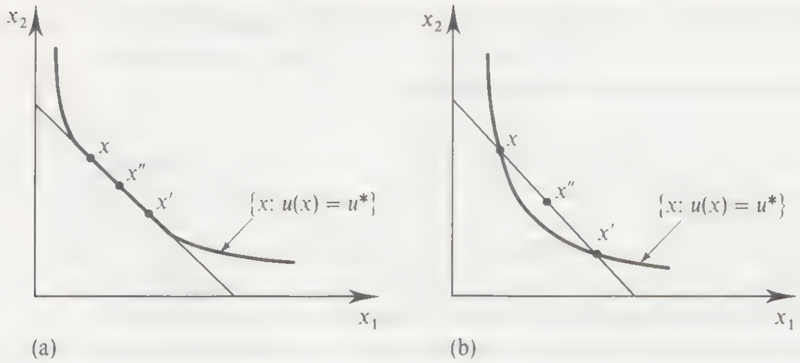


Figure 3.D.3
(a) Convexity of preferences implies convexity of $x(p, w)$.
(b) Strict convexity of preferences implies that $x(p, w)$ is single-valued.

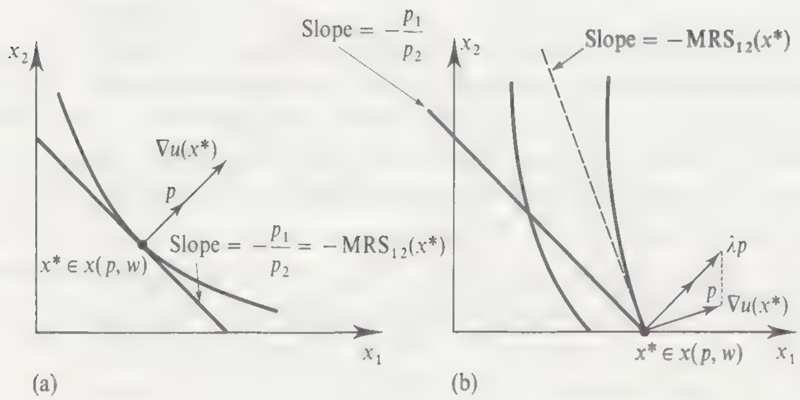


Figure 3.D.4
(a) Interior solution.
(b) Boundary solution.

If $u(\cdot)$ is continuously differentiable, an optimal consumption bundle $x^* \in x(p, w)$ can be characterized in a very useful manner by means of first-order conditions. The *Kuhn–Tucker (necessary) conditions* (see Section M.K of the Mathematical Appendix) say that if $x^* \in x(p, w)$ is a solution to the UMP, then there exists a *Lagrange multiplier* $\lambda \geq 0$ such that for all $\ell = 1, \ldots, L$:[8]

$$\frac{\partial u(x^*)}{\partial c_\ell} \leq \lambda p_\ell, \quad \text{with equality if } x_\ell^* > 0. \tag{3.D.1}$$

Equivalently, if we let $\nabla u(x) = [\partial u(x)/\partial x_1, \ldots, \partial u(x)/\partial x_L]$ denote the gradient vector of $u(\cdot)$ at x, we can write (3.D.1) in matrix notation as

$$\nabla u(x^*) \leq \lambda p \tag{3.D.2}$$

and

$$x^* \cdot [\nabla u(x^*) - \lambda p] = 0. \tag{3.D.3}$$

Thus, if we are at an interior optimum (i.e., if $x^* \gg 0$), we must have

$$\nabla u(x^*) = \lambda p. \tag{3.D.4}$$

Figure 3.D.4(a) depicts the first-order conditions for the case of an interior optimum when $L = 2$. Condition (3.D.4) tells us that at an interior optimum, the

8. To be fully rigorous, these Kuhn–Tucker necessary conditions are valid only if the constraint qualification condition holds (see Section M.K of the Mathematical Appendix). In the UMP, this is always so. Whenever we use Kuhn–Tucker necessary conditions without mentioning the constraint qualification condition, this requirement is met.

gradient vector of the consumer's utility function $\nabla u(x^*)$ must be proportional to the price vector p, as is shown in Figure 3.D.4(a). If $\nabla u(x^*) \gg 0$, this is equivalent to the requirement that for any two goods ℓ and k, we have

$$\frac{\partial u(x^*)/\partial x_\ell}{\partial u(x^*)/\partial x_k} = \frac{p_\ell}{p_k}. \tag{3.D.5}$$

The expression on the left of (3.D.5) is the *marginal rate of substitution of good ℓ for good k at x^*, $MRS_{\ell k}(x^*)$*; it tells us the amount of good k that the consumer must be given to compensate her for a one-unit marginal reduction in her consumption of good ℓ.[9] In the case where $L = 2$, the slope of the consumer's indifference set at x^* is precisely $-MRS_{12}(x^*)$. Condition (3.D.5) tells us that at an interior optimum, the consumer's marginal rate of substitution between any two goods must be equal to their price ratio, the marginal rate of exchange between them, as depicted in Figure 3.D.4(a). Were this not the case, the consumer could do better by marginally changing her consumption. For example, if $[\partial u(x^*)/\partial x_\ell]/[\partial u(x^*)/\partial x_k] > (p_\ell/p_k)$, then an increase in the consumption of good ℓ of dx_ℓ, combined with a decrease in good k's consumption equal to $(p_\ell/p_k)\,dx_\ell$, would be feasible and would yield a utility change of $[\partial u(x^*)/\partial x_\ell]\,dx_\ell - [\partial u(x^*)/\partial x_k](p_\ell/p_k)\,dx_\ell > 0$.

Figure 3.D.4(b) depicts the first-order conditions for the case of $L = 2$ when the consumer's optimal bundle x^* lies on the boundary of the consumption set (we have $x_2^* = 0$ there). In this case, the gradient vector need not be proportional to the price vector. In particular, the first-order conditions tell us that $\partial u_\ell(x^*)/\partial x_\ell \le \lambda p_\ell$ for those ℓ with $x_\ell^* = 0$ and $\partial u_\ell(x^*)/\partial x_\ell = \lambda p_\ell$ for those ℓ with $x_\ell^* > 0$. Thus, in the figure, we see that $MRS_{12}(x^*) > p_1/p_2$. In contrast with the case of an interior optimum, an inequality between the marginal rate of substitution and the price ratio can arise at a boundary optimum because the consumer is unable to reduce her consumption of good 2 (and correspondingly increase her consumption of good 1) any further.

The Lagrange multiplier λ in the first-order conditions (3.D.2) and (3.D.3) gives the marginal, or shadow, value of relaxing the constraint in the UMP (this is a general property of Lagrange multipliers; see Sections M.K and M.L of the Mathematical Appendix). It therefore equals the consumer's marginal utility value of wealth at the optimum. To see this directly, consider for simplicity the case where $x(p, w)$ is a differentiable function and $x(p, w) \gg 0$. By the chain rule, the change in utility from a marginal increase in w is given by $\nabla u(x(p, w))\cdot D_w x(p, w)$, where $D_w x(p, w) = [\partial x_1(p, w)/\partial w, \dots, \partial x_L(p, w)/\partial w]$. Substituting for $\nabla u(x(p, w))$ from condition (3.D.4), we get

$$\nabla u(x(p, w))\cdot D_w x(p, w) = \lambda p\cdot D_w x(p, w) = \lambda,$$

where the last equality follows because $p\cdot x(p, w) = w$ holds for all w (Walras' law) and therefore $p\cdot D_w x(p, w) = 1$. Thus, the marginal change in utility arising from

9. Note that if utility is unchanged with differential changes in x_ℓ and x_k, dx_ℓ and dx_k, then $[\partial u(x)/\partial x_\ell]\,dx_\ell + [\partial u(x)/\partial x_k]\,dx_k = 0$. Thus, when x_ℓ falls by amount $dx_\ell < 0$, the increase required in x_k to keep utility unchanged is precisely $dx_k = MRS_{\ell k}(x^*)(-dx_\ell)$.

a marginal increase in wealth—the consumer's *marginal utility of wealth*—is precisely λ.[10]

We have seen that conditions (3.D.2) and (3.D.3) must necessarily be satisfied by any $x^* \in x(p, w)$. When, on the other hand, does satisfaction of these first-order conditions by some bundle x imply that x is a solution to the UMP? That is, when are the first-order conditions *sufficient* to establish that x is a solution? If $u(\cdot)$ is quasiconcave and monotone and has $\nabla u(x) \neq 0$ for all $x \in \mathbb{R}^L_+$, then the Kuhn–Tucker first-order conditions are indeed sufficient (see Section M.K of the Mathematical Appendix). What if $u(\cdot)$ is not quasiconcave? In that case, if $u(\cdot)$ is locally quasiconcave at x^*, and if x^* satisfies the first-order conditions, then x^* is a local maximum. Local quasiconcavity can be verified by means of a determinant test on the *bordered Hessian matrix* of $u(\cdot)$ at x^*. (For more on this, see Sections M.C and M.D of the Mathematical Appendix.)

Example 3.D.1 illustrates the use of the first-order conditions in deriving the consumer's optimal consumption bundle.

Example 3.D.1: *The Demand Function Derived from the Cobb–Douglas Utility Function.* A *Cobb–Douglas* utility function for $L = 2$ is given by $u(x_1, x_2) = kx_1^{\alpha}x_2^{1-\alpha}$ for some $\alpha \in (0, 1)$ and $k > 0$. It is increasing at all $(x_1, x_2) \gg 0$ and is homogeneous of degree one. For our analysis, it turns out to be easier to use the increasing transformation $\alpha \ln x_1 + (1 - \alpha) \ln x_2$, a strictly concave function, as our utility function. With this choice, the UMP can be stated as

$$\underset{x_1, x_2}{\text{Max}} \quad \alpha \ln x_1 + (1 - \alpha) \ln x_2 \qquad (3.D.6)$$
$$\text{s.t. } p_1 x_1 + p_2 x_2 = w.$$

[Note that since $u(\cdot)$ is increasing, the budget constraint will hold with strict equality at any solution.]

Since $\ln 0 = -\infty$, the optimal choice $(x_1(p, w), x_2(p, w))$ is strictly positive and must satisfy the first-order conditions (we write the consumption levels simply as x_1 and x_2 for notational convenience)

$$\frac{\alpha}{x_1} = \lambda p_1 \quad \Rightarrow \lambda = \frac{\alpha}{x_1 p_1} \qquad (3.D.7)$$

and

$$\frac{1 - \alpha}{x_2} = \lambda p_2 \quad \Rightarrow \lambda = \frac{1-\alpha}{x} \qquad (3.D.8)$$

for some $\lambda \geq 0$, and the budget constraint $p \cdot x(p, w) = w$. Conditions (3.D.7) and (3.D.8) imply that

$$p_1 x_1 = \frac{\alpha}{1 - \alpha} p_2 x_2$$

or, using the budget constraint,

$$p_1 x_1 = \frac{\alpha}{1 - \alpha}(w - p_1 x_1).$$

---

10. Note that if monotonicity of $u(\cdot)$ is strengthened slightly by requiring that $\nabla u(x) \geq 0$ *and* $\nabla u(x) \neq 0$ for all x, then condition (3.D.4) and $p \gg 0$ also imply that λ is strictly positive at any solution of the UMP.

Hence (including the arguments of x_1 and x_2 once again)

$$x_1(p,w) = \frac{\alpha w}{p_1},$$

and (using the budget constraint)

$$x_2(p,w) = \frac{(1-\alpha)w}{p_2}.$$

Note that with the Cobb–Douglas utility function, the expenditure on each commodity is a constant fraction of wealth for any price vector p [a share of α goes for the first commodity and a share of $(1-\alpha)$ goes for the second]. ■

Exercise 3.D.1: Verify the three properties of Proposition 3.D.2 for the Walrasian demand function generated by the Cobb–Douglas utility function.

For the analysis of demand responses to changes in prices and wealth, it is also very helpful if the consumer's Walrasian demand is suitably continuous and differentiable. Because the issues are somewhat more technical, we will discuss the conditions under which demand satisfies these properties in Appendix A to this chapter. We conclude there that both properties hold under fairly general conditions. Indeed, if preferences are continuous, strictly convex, and locally nonsatiated on the consumption set $\mathbb{R}^L_+$, then $x(p, w)$ (which is then a function) is *always* continuous at all $(p, w) \gg 0$.

The Indirect Utility Function

For each $(p, w) \gg 0$, the utility value of the UMP is denoted $v(p, w) \in \mathbb{R}$. It is equal to $u(x^*)$ for any $x^* \in x(p, w)$. The function $v(p, w)$ is called the *indirect utility function* and often proves to be a very useful analytic tool. Proposition 3.D.3 identifies its basic properties.

Proposition 3.D.3: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. The indirect utility function $v(p, w)$ is

(i) Homogeneous of degree zero.
(ii) Strictly increasing in w and nonincreasing in p_ℓ for any ℓ.
(iii) Quasiconvex; that is, the set $\{(p, w): v(p, w) \leq \bar{v}\}$ is convex for any $\bar{v}$.[11]
(iv) Continuous in p and w.

Proof: Except for quasiconvexity and continuity all the properties follow readily from our previous discussion. We forgo the proof of continuity here but note that, when preferences are strictly convex, it follows from the fact that $x(p, w)$ and $u(x)$ are continuous functions because $v(p, w) = u(x(p, w))$ [recall that the continuity of $x(p, w)$ is established in Appendix A of this chapter].

To see that $v(p,w)$ is quasiconvex, suppose that $v(p, w) \leq \bar{v}$ and $v(p', w') \leq \bar{v}$. For any $\alpha \in [0, 1]$, consider then the price–wealth pair $(p'', w'') = (\alpha p + (1-\alpha)p', \alpha w + (1-\alpha)w')$.

11. Note that property (iii) says that $v(p, w)$ is quasi*convex*, *not* quasi*concave*. Observe also that property (iii) does not require for its validity that $u(\cdot)$ be quasiconcave.

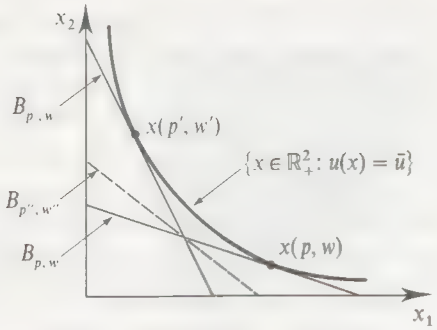


Figure 3.D.5
The indirect utility function $v(p, w)$ is quasiconvex.

To establish quasiconvexity, we want to show that $v(p'', w'') \leq \bar{v}$. Thus, we show that for any x with $p'' \cdot x \leq w''$, we must have $u(x) \leq \bar{v}$. Note, first, that if $p'' \cdot x \leq w''$, then,

$$\alpha p \cdot x + (1 - \alpha) p' \cdot x \leq \alpha w + (1 - \alpha) w'.$$

Hence, either $p \cdot x \leq w$ or $p' \cdot x \leq w'$ (or both). If the former inequality holds, then $u(x) \leq v(p, w) \leq \bar{v}$, and we have established the result. If the latter holds, then $u(x) \leq v(p', w') \leq \bar{v}$, and the same conclusion follows. ■

The quasiconvexity of $v(p, w)$ can be verified graphically in Figure 3.D.5 for the case where $L = 2$. There, the budget sets for price–wealth pairs (p, w) and (p', w') generate the same maximized utility value $\bar{u}$. The budget line corresponding to $(p'', w'') = (\alpha p + (1 - \alpha) p', \alpha w + (1 - \alpha) w')$ is depicted as a dashed line in Figure 3.D.5. Because (p'', w'') is a convex combination of (p, w) and (p', w'), its budget line lies between the budget lines for these two price–wealth pairs. As can be seen in the figure, the attainable utility under (p'', w'') is necessarily no greater than $\bar{u}$.

Note that the indirect utility function depends on the utility representation chosen. In particular, if $v(p, w)$ is the indirect utility function when the consumer's utility function is $u(\cdot)$, then the indirect utility function corresponding to utility representation $\tilde{u}(x) = f(u(x))$ is $\tilde{v}(p, w) = f(v(p, w))$.

Example 3.D.2: Suppose that we have the utility function $u(x_1, x_2) = \alpha \ln x_1 + (1 - \alpha) \ln x_2$. Then, substituting $x_1(p, w)$ and $x_2(p, w)$ from Example 3.D.1, into $u(x)$ we have

$$\begin{aligned} v(p, w) &= u(x(p, w)) \\ &= [\alpha \ln \alpha + (1 - \alpha) \ln (1 - \alpha)] + \ln w - \alpha \ln p_1 - (1 - \alpha) \ln p_2. \end{aligned}$$

Exercise 3.D.2: Verify the four properties of Proposition 3.D.3 for the indirect utility function derived in Example 3.D.2.

3.E The Expenditure Minimization Problem

In this section, we study the following *expenditure minimization problem* (EMP) for $p \gg 0$ and $u > u(0)$:[12]

12. Utility $u(0)$ is the utility from consuming the consumption bundle $x = (0, 0, \ldots, 0)$. The restriction to $u > u(0)$ rules out only uninteresting situations.

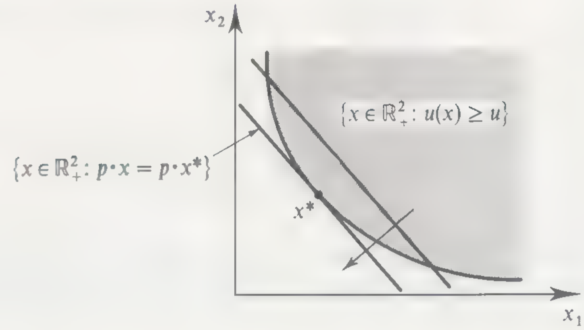


Figure 3.E.1
The expenditure minimization problem (EMP).

$$\underset{x \geq 0}{\text{Min}} \quad p \cdot x \qquad \text{(EMP)}$$
$$\text{s.t. } u(x) \geq u.$$

Whereas the UMP computes the maximal level of utility that can be obtained given wealth w, the EMP computes the minimal level of wealth required to reach utility level u. The EMP is the "dual" problem to the UMP. It captures the same aim of efficient use of the consumer's purchasing power while reversing the roles of objective function and constraint.[13]

Throughout this section, we assume that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $\mathbb{R}^L_+$.

The EMP is illustrated in Figure 3.E.1. The optimal consumption bundle x^* is the least costly bundle that still allows the consumer to achieve the utility level u. Geometrically, it is the point in the set $\{x \in \mathbb{R}^L_+ : u(x) \geq u\}$ that lies on the lowest possible budget line associated with the price vector p.

Proposition 3.E.1 describes the formal relationship between EMP and the UMP.

Proposition 3.E.1: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$ and that the price vector is $p \gg 0$. We have

(i) If x^* is optimal in the UMP when wealth is $w > 0$, then x^* is optimal in the EMP when the required utility level is $u(x^*)$. Moreover, the minimized expenditure level in this EMP is exactly w.

(ii) If x^* is optimal in the EMP when the required utility level is $u > u(0)$, then x^* is optimal in the UMP when wealth is $p \cdot x^*$. Moreover, the maximized utility level in this UMP is exactly u.

Proof: (i) Suppose that x^* is not optimal in the EMP with required utility level $u(x^*)$. Then there exists an x' such that $u(x') \geq u(x^*)$ and $p \cdot x' < p \cdot x^* \leq w$. By local nonsatiation, we can find an x'' very close to x' such that $u(x'') > u(x')$ and $p \cdot x'' < w$. But this implies that $x'' \in B_{p,w}$ and $u(x'') > u(x^*)$, contradicting the optimality of x^* in the UMP. Thus, x^* must be optimal in the EMP when the required utility level

13. The term "dual" is meant to be suggestive. It is usually applied to pairs of problems and concepts that are formally similar except that the role of quantities and prices, and/or maximization and minimization, and/or objective function and constraint, have been reversed.

is $u(x^*)$, and the minimized expenditure level is therefore $p \cdot x^*$. Finally, since x^* solves the UMP when wealth is w, by Walras' law we have $p \cdot x^* = w$.

(ii) Since $u > u(0)$, we must have $x^* \neq 0$. Hence, $p \cdot x^* > 0$. Suppose that x^* is not optimal in the UMP when wealth is $p \cdot x^*$. Then there exists an x' such that $u(x') > u(x^*)$ and $p \cdot x' \leq p \cdot x^*$. Consider a bundle $x'' = \alpha x'$ where $\alpha \in (0,1)$ (x'' is a "scaled-down" version of x'). By continuity of $u(\cdot)$, if α is close enough to 1, then we will have $u(x'') > u(x^*)$ and $p \cdot x'' < p \cdot x^*$. But this contradicts the optimality of x^* in the EMP. Thus, x^* must be optimal in the UMP when wealth is $p \cdot x^*$, and the maximized utility level is therefore $u(x^*)$. In Proposition 3.E.3(ii), we will show that if x^* solves the EMP when the required utility level is u, then $u(x^*) = u$. ■

As with the UMP, when $p \gg 0$ a solution to the EMP exists under very general conditions. The constraint set merely needs to be nonempty; that is, $u(\cdot)$ must attain values at least as large as u for *some* x (see Exercise 3.E.3). From now on, we assume that this is so; for example, this condition will be satisfied for any $u > u(0)$ if $u(\cdot)$ is unbounded above.

We now proceed to study the optimal consumption vector and the value function of the EMP. We consider the value function first.

The Expenditure Function

Given prices $p \gg 0$ and required utility level $u > u(0)$, the value of the EMP is denoted $e(p, u)$. The function $e(p, u)$ is called the *expenditure function*. Its value for any (p, u) is simply $p \cdot x^*$, where x^* is any solution to the EMP. The result in Proposition 3.E.2 describes the basic properties of the expenditure function. It parallels Proposition 3.D.3's characterization of the properties of the indirect utility function for the UMP.

Proposition 3.E.2: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. The expenditure function $e(p, u)$ is

(i) Homogeneous of degree one in p.
(ii) Strictly increasing in u and nondecreasing in p_ℓ for any ℓ.
(iii) Concave in p.
(iv) Continuous in p and u.

Proof: We prove only properties (i), (ii), and (iii).

(i) The constraint set of the EMP is unchanged when prices change. Thus, for any scalar $\alpha > 0$, minimizing $(\alpha p) \cdot x$ on this set leads to the same optimal consumption bundles as minimizing $p \cdot x$. Letting x^* be optimal in both circumstances, we have $e(\alpha p, u) = \alpha p \cdot x^* = \alpha e(p, u)$.

(ii) Suppose that $e(p, u)$ were not strictly increasing in u, and let x' and x'' denote optimal consumption bundles for required utility levels u' and u'', respectively, where $u'' > u'$ and $p \cdot x' \geq p \cdot x'' > 0$. Consider a bundle $\tilde{x} = \alpha x''$, where $\alpha \in (0, 1)$. By continuity of $u(\cdot)$, there exists an α close enough to 1 such that $u(\tilde{x}) > u'$ and $p \cdot x' > p \cdot \tilde{x}$. But this contradicts x' being optimal in the EMP with required utility level u'.

To show that $e(p, u)$ is nondecreasing in p_ℓ, suppose that price vectors p'' and p' have $p''_\ell \geq p'_\ell$ and $p''_k = p'_k$ for all $k \neq \ell$. Let x'' be an optimizing vector in the EMP for prices p''. Then $e(p'', u) = p'' \cdot x'' \geq p' \cdot x'' \geq e(p', u)$, where the latter inequality follows from the definition of $e(p', u)$.

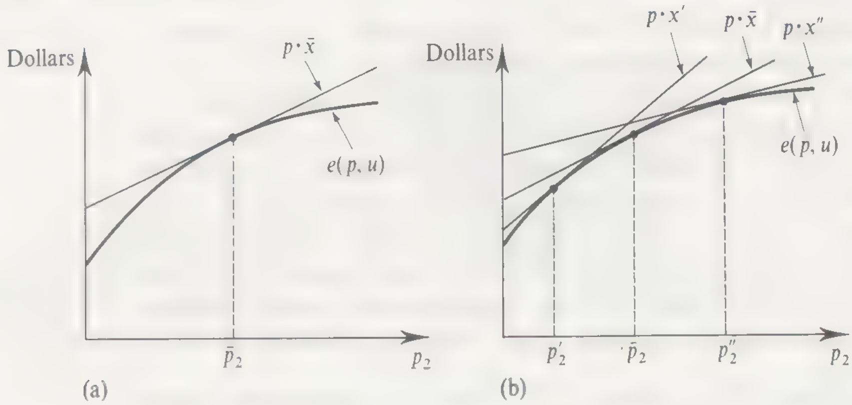


Figure 3.E.2
The concavity in p of the expenditure function.

(iii) For concavity, fix a required utility level $\bar{u}$, and let $p'' = \alpha p + (1 - \alpha)p'$ for $\alpha \in [0, 1]$. Suppose that x'' is an optimal bundle in the EMP when prices are p''. If so,

$$\begin{aligned} e(p'', \bar{u}) &= p'' \cdot x'' \\ &= \alpha p \cdot x'' + (1 - \alpha)p' \cdot x'' \\ &\geq \alpha e(p, \bar{u}) + (1 - \alpha)e(p', \bar{u}), \end{aligned}$$

where the last inequality follows because $u(x'') \geq \bar{u}$ and the definition of the expenditure function imply that $p \cdot x'' \geq e(p, \bar{u})$ and $p' \cdot x'' \geq e(p', \bar{u})$. ■

The concavity of $e(p, \bar{u})$ in p for given $\bar{u}$, which is a very important property, is actually fairly intuitive. Suppose that we initially have prices $\bar{p}$ and that $\bar{x}$ is an optimal consumption vector at these prices in the EMP. If prices change but we do not let the consumer change her consumption levels from $\bar{x}$, then the resulting expenditure will be $p \cdot \bar{x}$, which is a *linear* expression in p. But when the consumer can adjust her consumption, as in the EMP, her minimized expenditure level can be no greater than this amount. Hence, as illustrated in Figure 3.E.2(a), where we keep p_1 fixed and vary p_2, the graph of $e(p, \bar{u})$ lies below the graph of the linear function $p \cdot \bar{x}$ at all $p \neq \bar{p}$ and touches it at $\bar{p}$. This amounts to concavity because a similar relation to a linear function must hold at each point of the graph of $e(\cdot, u)$; see Figure 3.E.2(b).

Proposition 3.E.1 allows us to make an important connection between the expenditure function and the indirect utility function developed in Section 3.D. In particular, for any $p \gg 0$, $w > 0$, and $u > u(0)$ we have

$$e(p, v(p, w)) = w \qquad \text{and} \qquad v(p, e(p, u)) = u. \tag{3.E.1}$$

These conditions imply that for a fixed price vector $\bar{p}$, $e(\bar{p}, \cdot)$ and $v(\bar{p}, \cdot)$ are inverses to one another (see Exercise 3.E.8). In fact, in Exercise 3.E.9, you are asked to show that by using the relations in (3.E.1), Proposition 3.E.2 can be directly derived from Proposition 3.D.3, and vice versa. That is, there is a direct correspondence between the properties of the expenditure function and the indirect utility function. They both capture the same underlying features of the consumer's choice problem.

The Hicksian (or Compensated) Demand Function

The set of optimal commodity vectors in the EMP is denoted $h(p, u) \subset \mathbb{R}^L_+$ and is known as the *Hicksian*, or *compensated, demand correspondence*, or *function* if

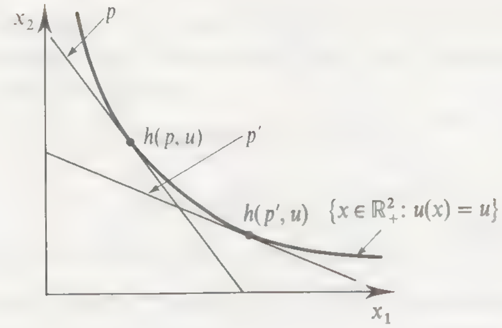


Figure 3.E.3
The Hicksian (or compensated) demand function.

single-valued. (The reason for the term "compensated demand" will be explained below.) Figure 3.E.3 depicts the solution set $h(p, u)$ for two different price vectors p and p'.

Three basic properties of Hicksian demand are given in Proposition 3.E.3, which parallels Proposition 3.D.2 for Walrasian demand.

Proposition 3.E.3: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. Then for any $p \gg 0$, the Hicksian demand correspondence $h(p, u)$ possesses the following properties:

(i) *Homogeneity of degree zero in* p: $h(\alpha p, u) = h(p, u)$ for any p, u and $\alpha > 0$.
(ii) *No excess utility:* For any $x \in h(p, u)$, $u(x) = u$.
(iii) *Convexity/uniqueness:* If $\succsim$ is convex, then $h(p, u)$ is a convex set; and if $\succsim$ is *strictly* convex, so that $u(\cdot)$ is strictly quasiconcave, then there is a unique element in $h(p, u)$.

Proof: (i) Homogeneity of degree zero in p follows because the optimal vector when minimizing $p \cdot x$ subject to $u(x) \geq u$ is the same as that for minimizing $\alpha p \cdot x$ subject to this same constraint, for any scalar $\alpha > 0$.

(ii) This property follows from continuity of $u(\cdot)$. Suppose there exists an $x \in h(p, u)$ such that $u(x) > u$. Consider a bundle $x' = \alpha x$, where $\alpha \in (0, 1)$. By continuity, for α close enough to 1, $u(x') \geq u$ and $p \cdot x' < p \cdot x$, contradicting x being optimal in the EMP with required utility level u.

(iii) The proof of property (iii) parallels that for property (iii) of Proposition 3.D.2 and is left as an exercise (Exercise 3.E.4). ■

As in the UMP, when $u(\cdot)$ is differentiable, the optimal consumption bundle in the EMP can be characterized using first-order conditions. As would be expected given Proposition 3.E.1, these first-order conditions bear a close similarity to those of the UMP. Exercise 3.E.1 asks you to explore this relationship.

Exercise 3.E.1: Assume that $u(\cdot)$ is differentiable. Show that the first-order conditions for the EMP are

$$p \geq \lambda \nabla u(x^*) \tag{3.E.2}$$

and

$$x^* \cdot [p - \lambda \nabla u(x^*)] = 0, \tag{3.E.3}$$

for some $\lambda \geq 0$. Compare this with the first-order conditions of the UMP.

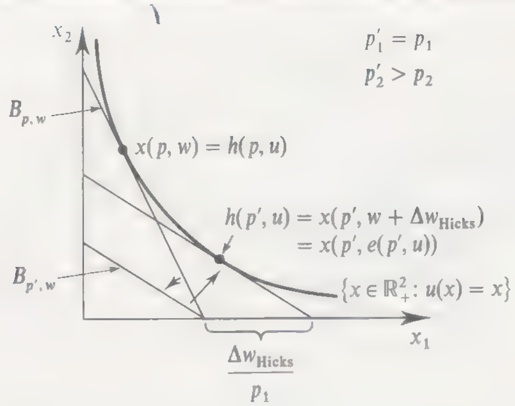


Figure 3.E.4
Hicksian wealth compensation.

We will not discuss the continuity and differentiability properties of the Hicksian demand correspondence. With minimal qualifications, they are the same as for the Walrasian demand correspondence, which we discuss in some detail in Appendix A.

Using Proposition 3.E.1, we can relate the Hicksian and Walrasian demand correspondences as follows:

$$h(p, u) = x(p, e(p, u)) \quad \text{and} \quad x(p, w) = h(p, v(p, w)). \tag{3.E.4}$$

The first of these relations explains the use of the term *compensated demand correspondence* to describe $h(p, u)$: As prices vary, $h(p, u)$ gives precisely the level of demand that would arise if the consumer's wealth were simultaneously adjusted to keep her utility level at u. This type of wealth compensation, which is depicted in Figure 3.E.4, is known as *Hicksian wealth compensation.* In Figure 3.E.4, the consumer's initial situation is the price–wealth pair (p, w), and prices then change to p', where $p'_1 = p_1$ and $p'_2 > p_2$. The Hicksian wealth compensation is the amount $\Delta w_{\text{Hicks}} = e(p', u) - w$. Thus, the demand function $h(p, u)$ keeps the consumer's utility level fixed as prices change, in contrast with the Walrasian demand function, which keeps money wealth fixed but allows utility to vary.

As with the value functions of the EMP and UMP, the relations in (3.E.4) allow us to develop a tight linkage between the properties of the Hicksian demand correspondence $h(p, u)$ and the Walrasian demand correspondence $x(p, w)$. In particular, in Exercise 3.E.10, you are asked to use the relations in (3.E.4) to derive the properties of each correspondence as a direct consequence of those of the other.

Hicksian Demand and the Compensated Law of Demand

An important property of Hicksian demand is that it satisfies the *compensated law of demand*: Demand and price move in opposite directions for price changes that are accompanied by Hicksian wealth compensation. In Proposition 3.E.4, we prove this fact for the case of single-valued Hicksian demand.

Proposition 3.E.4: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated preference relation $\succsim$ and that $h(p, u)$ consists of a single element for all $p \gg 0$. Then the Hicksian demand function $h(p, u)$ satisfies the compensated law of demand: For all p' and p'',

$$(p'' - p') \cdot [h(p'', u) - h(p', u)] \leq 0. \tag{3.E.5}$$

Proof: For any $p \gg 0$, consumption bundle $h(p, u)$ is optimal in the EMP, and so it achieves a lower expenditure at prices p than any other bundle that offers a utility level of at least u. Therefore, we have

$$p'' \cdot h(p'', u) \leq p'' \cdot h(p', u)$$

and

$$p' \cdot h(p'', u) \geq p' \cdot h(p', u).$$

Subtracting these two inequalities yields the results. ■

One immediate implication of Proposition 3.E.4 is that for compensated demand, own-price effects are nonpositive. In particular, if only p_ℓ changes, Proposition 3.E.4 implies that $(p''_\ell - p'_\ell)[h_\ell(p'', u) - h_\ell(p', u)] \leq 0$. The comparable statement is *not* true for Walrasian demand. Walrasian demand need not satisfy the law of demand. For example, the demand for a good can decrease when its price falls. See Section 2.E for a discussion of Giffen goods and Figure 2.F.5 (along with the discussion of that figure in Section 2.F) for a diagrammatic example.

Example 3.E.1: *Hicksian Demand and Expenditure Functions for the Cobb–Douglas Utility Function.* Suppose that the consumer has the Cobb–Douglas utility function over the two goods given in Example 3.D.1. That is, $u(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$. By deriving the first-order conditions for the EMP (see Exercise 3.E.1), and substituting from the constraint $u(h_1(p, u), h_2(p, u)) = u$, we obtain the Hicksian demand functions

$$h_1(p, u) = \left[\frac{\alpha p_2}{(1-\alpha)p_1}\right]^{1-\alpha} u$$

and

$$h_2(p, u) = \left[\frac{(1-\alpha)p_1}{\alpha p_2}\right]^{\alpha} u.$$

Calculating $e(p, u) = p \cdot h(p, u)$ yields

$$e(p, u) = [\alpha^{-\alpha}(1-\alpha)^{\alpha-1}] p_1^\alpha p_2^{1-\alpha} u. \quad ■$$

Exercise 3.E.2: Verify the properties listed in Propositions 3.E.2 and 3.E.3 for the Hicksian demand and expenditure functions of the Cobb–Douglas utility function.

Here and in the preceding section, we have derived several basic properties of the Walrasian and Hicksian demand functions, the indirect utility function, and the expenditure function. We investigate these concepts further in Section 3.G. First, however, in Section 3.F, which is meant as optional, we offer an introductory discussion of the mathematics underlying the theory of duality. The material covered in Section 3.F provides a better understanding of the essential connections between the UMP and the EMP. We emphasize, however, that this section is not a prerequisite for the study of the remaining sections of this chapter.

3.F Duality: A Mathematical Introduction

This section constitutes a mathematical detour. It focuses on some aspects of the theory of convex sets and functions.

Recall that a set $K \subset \mathbb{R}^L$ is convex if $\alpha x + (1-\alpha)z \in K$ whenever $x, z \in K$ and $\alpha \in [0, 1]$. Note that the intersection of two convex sets is a convex set.

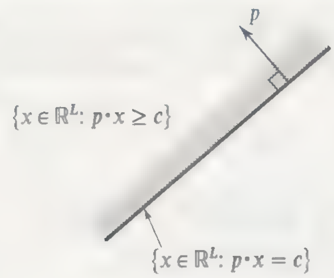


Figure 3.F.1
A half-space and a hyperplane.

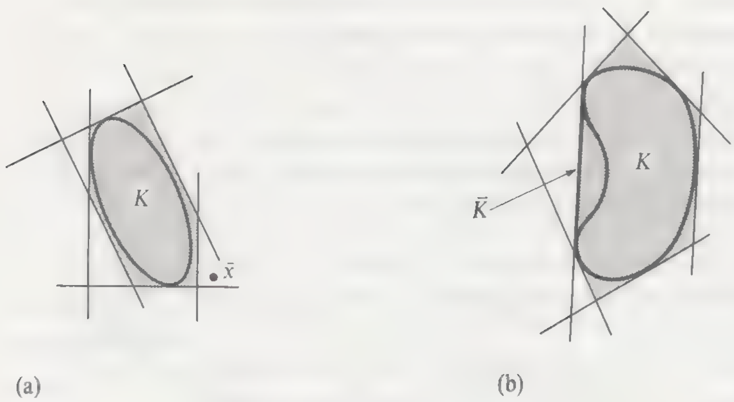


Figure 3.F.2
A closed set is convex if and only if it equals the intersection of the half-spaces that contain it.
(a) Convex K.
(b) Nonconvex K.

A *half-space* is a set of the form $\{x \in \mathbb{R}^L: p \cdot x \geq c\}$ for some $p \in \mathbb{R}^L$, $p \neq 0$, called the *normal vector* to the half-space, and some $c \in \mathbb{R}$. Its boundary $\{x \in \mathbb{R}^L: p \cdot x = c\}$ is called a *hyperplane*. The term *normal* comes from the fact that whenever $p \cdot x = p \cdot x' = c$, we have $p \cdot (x - x') = 0$, and so p is orthogonal (i.e., perpendicular, or normal) to the hyperplane (see Figure 3.F.1). Note that both half-spaces and hyperplanes are convex sets.

Suppose now that $K \subset \mathbb{R}^L$ is a convex set that is also closed (i.e., it includes its boundary points), and consider any point $\bar{x} \notin K$ outside of this set. A fundamental theorem of convexity theory, the *separating hyperplane theorem*, tells us that there is a half-space containing K and excluding $\bar{x}$ (see Section M.G of the Mathematical Appendix). That is, there is a $p \in \mathbb{R}^L$ and a $c \in \mathbb{R}$ such that $p \cdot \bar{x} < c \leq p \cdot x$ for all $x \in K$. The basic idea behind duality theory is the fact that a closed, convex set can equivalently ("dually") be described as the intersection of the half-spaces that contain it; this is illustrated in Figure 3.F.2(a). Because any $\bar{x} \notin K$ is excluded by some half-space that contains K, as we draw such half-spaces for more and more points $\bar{x} \notin K$, their intersection (the shaded area in the figure) becomes equal to K.

More generally, if the set K is not convex, the intersection of the half-spaces that contain K is the smallest closed, convex set that contains K, known as the *closed, convex hull* of K. Figure 3.F.2(b) illustrates a case where the set K is nonconvex; in the figure, the closed convex hull of K is $\bar{K}$.

Given any closed (but not necessarily convex) set $K \subset \mathbb{R}^L$ and a vector $p \in \mathbb{R}^L$, we can define the *support function* of K.

Definition 3.F.1: For any nonempty closed set $K \subset \mathbb{R}^L$, the *support function* of K is defined for any $p \in \mathbb{R}^L$ to be

$$\mu_K(p) = \text{Infimum}\,\{p \cdot x: x \in K\}.$$

The *infimum* of a set of numbers, as used in Definition 3.F.1, is a generalized version of the set's minimum value. In particular, it allows for situations in which no minimum exists because although points in the set can be found that come arbitrarily close to some lower bound value, no point in the set actually attains that value. For example, consider a strictly positive function $f(x)$ that approaches zero asymptotically as x increases. The minimum of this function does not exist, but its infimum is zero. The formulation also allows $\mu_K(p)$ to take the value $-\infty$ when points in K can be found that make the value of $p \cdot x$ unboundedly negative.

When K is convex, the function $\mu_K(\cdot)$ provides an alternative ("dual") description of K because we can reconstruct K from knowledge of $\mu_K(\cdot)$. In particular, for every p, $\{x \in \mathbb{R}^L: p \cdot x \geq \mu_K(p)\}$ is a half-space that contains K. In addition, as we discussed above, if $x \notin K$, then $p \cdot x < \mu_K(p)$ for some p. Thus, the intersection of the half-spaces generated by all possible values of p is precisely K; that is,

$$K = \{x \in \mathbb{R}^L: p \cdot x \geq \mu_K(p) \text{ for every } p\}.$$

By the same logic, if K is not convex, then $\{x \in \mathbb{R}^L: p \cdot x \geq \mu_K(p) \text{ for every } p\}$ is the smallest closed, convex set containing K.

The function $\mu_K(\cdot)$ is homogeneous of degree one. More interestingly, *it is concave.* To see this, consider $p'' = \alpha p + (1 - \alpha)p'$ for $\alpha \in [0,1]$. To make things simple, suppose that the infimum is in fact attained, so that there is a $z \in K$ such that $\mu_K(p'') = p'' \cdot z$. Then, because

$$\begin{aligned} \mu_K(p'') &= \alpha p \cdot z + (1 - \alpha)p' \cdot z \\ &\geq \alpha \mu_K(p) + (1 - \alpha)\mu_K(p'). \end{aligned}$$

we conclude that $\mu_K(\cdot)$ is concave.

The concavity of $\mu_K(\cdot)$ can also be seen geometrically. Figure 3.F.3 depicts the value of the function $\phi_x(p) = p \cdot x$, for various choices of $x \in K$, as a function of p_2 (with p_1 fixed at $\bar{p}_1$). For each x, the function $\phi_x(\cdot)$ is a linear function of p_2. Also shown in the figure is $\mu_K(\cdot)$. For each level of p_2, $\mu_K(\bar{p}_1, p_2)$ is equal to the minimum value (technically, the infimum) of the various linear functions $\phi_x(\cdot)$ at $p = (\bar{p}_1, p_2)$; that is, $\mu_K(\bar{p}_1, p_2) = \text{Min}\,\{\phi_x(\bar{p}_1, p_2): x \in K\}$. For example, when $p_2 = \bar{p}_2$, $\mu_K(\bar{p}_1, \bar{p}_2) = \phi_{\bar{x}}(\bar{p}_1, \bar{p}_2) \leq \phi_x(\bar{p}_1, \bar{p}_2)$ for all $x \in K$. As can be seen in the figure, $\mu_K(\cdot)$ is therefore the "lower envelope" of the functions $\phi_x(\cdot)$. As the infimum of a family of linear functions, $\mu_K(\cdot)$ is concave.

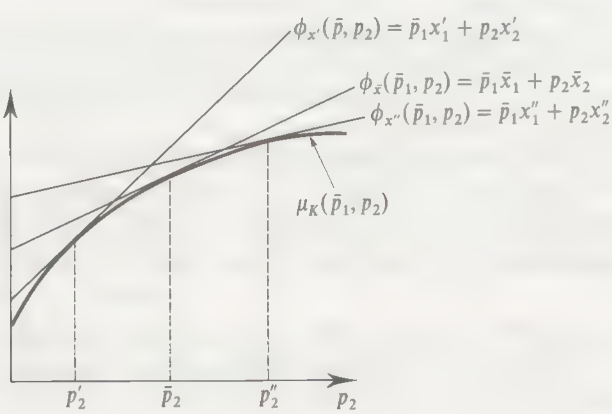


Figure 3.F.3
The support function $\mu_K(p)$ is concave.

Proposition 3.F.1, the *duality theorem*, gives the central result of the mathematical theory. Its use is pervasive in economics.

Proposition 3.F.1: (*The Duality Theorem*). Let K be a nonempty closed set, and let $\mu_K(\cdot)$ be its support function. Then there is a unique $\bar{x} \in K$ such that $\bar{p} \cdot \bar{x} = \mu_K(\bar{p})$ if and only if $\mu_K(\cdot)$ is differentiable at $\bar{p}$. Moreover, in this case,

$$\nabla \mu_K(\bar{p}) = \bar{x}.$$

We will not give a complete proof of the theorem. Its most important conclusion is that if the minimizing vector $\bar{x}$ for the vector $\bar{p}$ is unique, then the gradient of the support function at $\bar{p}$ is equal to $\bar{x}$. To understand this result, consider the linear function $\phi_{\bar{x}}(p) = p \cdot \bar{x}$. By the definition of $\bar{x}$, we know that $\mu_K(\bar{p}) = \phi_{\bar{x}}(\bar{p})$. Moreover, the derivatives of $\phi_{\bar{x}}(\cdot)$ at $\bar{p}$ satisfy $\nabla \phi_{\bar{x}}(p) = \bar{x}$. Therefore, the duality theorem tells us that as far as the first derivatives of $\mu_K(\cdot)$ are concerned, it is as if $\mu_K(\cdot)$ is linear in p; that is, the first derivatives of $\mu_K(\cdot)$ at $\bar{p}$ are exactly the same as those of the function $\phi_{\bar{x}}(p) = p \cdot \bar{x}$.

The logic behind this fact is relatively straightforward. Suppose that $\mu_K(\cdot)$ is differentiable at $\bar{p}$, and consider the function $\xi(p) = p \cdot \bar{x} - \mu_K(p)$, where $\bar{x} \in K$ and $\mu_K(\bar{p}) = \bar{p} \cdot \bar{x}$. By the definition of $\mu_K(\cdot)$, $\xi(p) = p \cdot \bar{x} - \mu_K(p) \geq 0$ for all p. We also know that $\xi(\bar{p}) = \bar{p} \cdot \bar{x} - \mu_K(\bar{p}) = 0$. So the function $\xi(\cdot)$ reaches a minimum at $p = \bar{p}$. As a result, its partial derivatives at $\bar{p}$ must all be zero. This implies the result: $\nabla \xi(\bar{p}) = \bar{x} - \nabla \mu_K(\bar{p}) = 0$.[14]

Recalling our discussion of the EMP in Section 3.E, we see that the expenditure function is precisely the support function of the set $\{x \in \mathbb{R}^L_+ : u(x) \geq u\}$. From our discussion of the support function, several of the properties of the expenditure function previously derived in Proposition 3.E.2, such as homogeneity of degree zero and concavity, immediately follow. In Section 3.G, we study the implications of the duality theorem for the theory of demand.

For a further discussion of duality theory and its applications, see Green and Heller (1981) and, for an advanced treatment, Diewert (1982). For an early application of duality to consumer theory, see McKenzie (1956–57).

The first part of the duality theorem says that $\mu_K(\cdot)$ is differentiable at $\bar{p}$ if and only if the minimizing vector at $\bar{p}$ is unique. If K is not strictly convex, then at some $\bar{p}$, the minimizing vector will not be unique and therefore $\mu_K(\cdot)$ will exhibit a kink at $\bar{p}$. Nevertheless, in a sense that can be made precise by means of the concept of directional derivatives, the gradient $\mu_K(\cdot)$ at this $\bar{p}$ is still equal to the minimizing set, which in this case is multivalued.

This is illustrated in Figure 3.F.4 for $L = 2$. In panel (a) of Figure 3.F.4, a strictly convex set K is depicted. For all p, its minimizing vector is unique. At $\bar{p} = (\frac{1}{2}, \frac{1}{2})$, it is $\bar{x} = (1, 1)$. Panel (b) of Figure 3.F.4 graphs $\mu_K(\frac{1}{2}, p_2)$ as a function of p_2. As can be seen, the function is concave and differentiable in p_2, with a slope of 1 (the value of $\bar{x}_2$) at $p_2 = \frac{1}{2}$.

In panel (a) of Figure 3.F.5, a convex but not strictly convex set K is depicted. At $\bar{p} = (\frac{1}{2}, \frac{1}{2})$, the entire segment $[x', x'']$ is the minimizing set. If $p_1 > p_2$, then x' is the minimizing vector and the value of the support function is $p_1 x'_1 + p_2 x'_2$, whereas if $p_1 < p_2$, then x'' is optimal and the value of the support function is $p_1 x''_1 + p_2 x''_2$. Panel (b) of Figure 3.F.5

14. Because $\bar{x} = \nabla \mu_K(\bar{p})$ for any minimizer $\bar{x}$ at $\bar{p}$, either $\bar{x}$ is unique or if it is not unique, then $\mu_K(\cdot)$ could not be differentiable at $\bar{p}$. Thus, $\mu_K(\cdot)$ is differentiable at $\bar{p}$ only if there is a unique minimizer at $\bar{p}$.

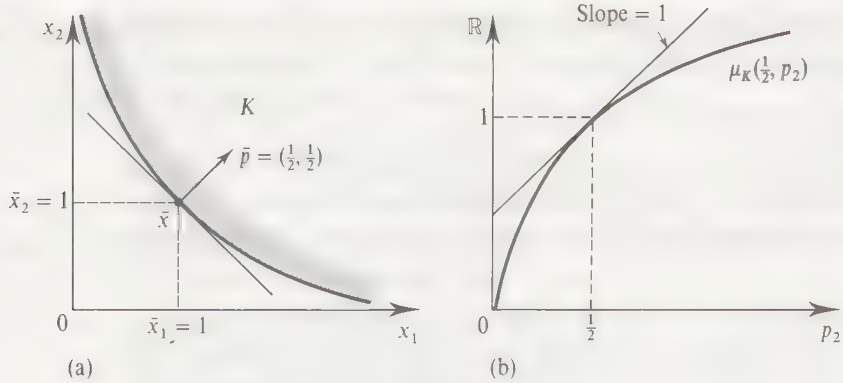


Figure 3.F.4
The duality theorem with a unique minimizing vector at $\bar{p}$. (a) The minimum vector. (b) The support function.

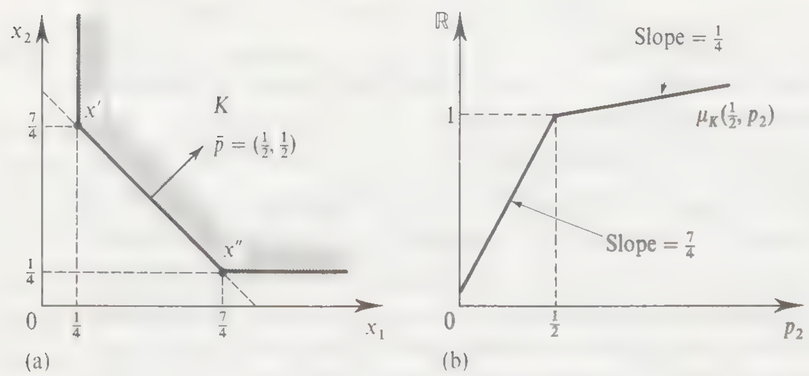


Figure 3.F.5
The duality theorem with a multivalued minimizing set at $\bar{p}$. (a) The minimum set. (b) The support function.

graphs $\mu_K(\frac{1}{2}, p_2)$ as a function of p_2. For $p_2 < \frac{1}{2}$, its slope is equal to $\frac{7}{4}$, the value of x_2'. For $p_2 > \frac{1}{2}$, its slope is $\frac{1}{4}$, the value of x_2''. There is a kink in the function at $\bar{p} = (\frac{1}{2}, \frac{1}{2})$, the price vector that has multiple minimizing vectors, with its left derivative with respect to p_2 equal to $\frac{7}{4}$ and its right derivative equal to $\frac{1}{4}$. Thus, the range of these directional derivatives at $\bar{p} = (\frac{1}{2}, \frac{1}{2})$ is equal to the range of x_2 in the minimizing vectors at that point.

G Relationships between Demand, Indirect Utility, and Expenditure Functions

We now continue our exploration of results flowing from the UMP and the EMP. The investigation in this section concerns three relationships: that between the Hicksian demand function and the expenditure function, that between the Hicksian and Walrasian demand functions, and that between the Walrasian demand function and the indirect utility function.

As before, we assume that $u(\cdot)$ is a continuous utility function representing the locally nonsatiated preferences $\succsim$ (defined on the consumption set $X = \mathbb{R}^L_+$), and we restrict attention to cases where $p \gg 0$. In addition, to keep matters simple, we assume

throughout that $\succsim$ is strictly convex, so that the Walrasian and Hicksian demands, $x(p, w)$ and $h(p, u)$, are single-valued.[15]

Hicksian Demand and the Expenditure Function

From knowledge of the Hicksian demand function, the expenditure function can readily be calculated as $e(p, u) = p \cdot h(p, u)$. The important result shown in Proposition 3.G.1 establishes a more significant link between the two concepts that runs in the opposite direction.

Proposition 3.G.1: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated and strictly convex preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. For all p and u, the Hicksian demand $h(p, u)$ is the derivative vector of the expenditure function with respect to prices:

$$h(p, u) = \nabla_p e(p, u). \tag{3.G.1}$$

That is, $h_\ell(p, u) = \partial e(p, u)/\partial p_\ell$ for all $\ell = 1, \ldots, L$.

Thus, given the expenditure function, we can calculate the consumer's Hicksian demand function simply by differentiating.

We provide three proofs of this important result.

Proof 1: (*Duality Theorem Argument*). The result is an immediate consequence of the duality theorem (Proposition 3.F.1). Since the expenditure function is precisely the support function for the set $K = \{x \in \mathbb{R}^L_+ : u(x) \geq u\}$, and since the optimizing vector associated with this support function is $h(p, u)$, Proposition 3.F.1 implies that $h(p, u) = \nabla_p e(p, u)$. Note that (3.G.1) helps us understand the use of the term "dual" in this context. In particular, just as the derivatives of the utility function $u(\cdot)$ with respect to quantities have a price interpretation (we have seen in Section 3.D that at an optimum they are equal to prices multiplied by a constant factor of proportionality), (3.G.1) tells us that the derivatives of the expenditure function $e(\cdot, u)$ with respect to prices have a quantity interpretation (they are equal to the Hicksian demands). ■

Proof 2: (*First-Order Conditions Argument*). For this argument, we focus for simplicity on the case where $h(p, u) \gg 0$, and we assume that $h(p, u)$ is differentiable at (p, u).

Using the chain rule, the change in expenditure can be written as

$$\begin{aligned} \nabla_p e(p, u) &= \nabla_p [p \cdot h(p, u)] \\ &= h(p, u) + [p \cdot D_p h(p, u)]^{\mathrm{T}}. \end{aligned} \tag{3.G.2}$$

Substituting from the first-order conditions for an interior solution to the EMP, $p = \lambda \nabla u(h(p, u))$, yields

$$\nabla_p e(p, u) = h(p, u) + \lambda [\nabla u(h(p, u)) \cdot D_p h(p, u)]^{\mathrm{T}}.$$

But since the constraint $u(h(p, u)) = u$ holds for all p in the EMP, we know that $\nabla u(h(p, u)) \cdot D_p h(p, u) = 0$, and so we have the result. ■

15. In fact, all the results of this section are local results that hold at all price vectors $\bar{p}$ with the property that for all p near $\bar{p}$, the optimal consumption vector in the UMP or EMP with price vector p is unique.

Proof 3: (*Envelope Theorem Argument*). Under the same simplifying assumptions used in Proof 2, we can directly appeal to the *envelope theorem*. Consider the value function $\phi(\alpha)$ of the constrained minimization problem

$$\begin{aligned} \underset{x}{\text{Min}} \quad & f(x, \alpha) \\ \text{s.t. } & g(x, \alpha) = 0. \end{aligned}$$

If $x^*(\alpha)$ is the (differentiable) solution to this problem as a function of the parameters $\alpha = (\alpha_1, \ldots, \alpha_M)$, then the envelope theorem tells us that at any $\bar{\alpha} = (\bar{\alpha}_1, \ldots, \bar{\alpha}_M)$ we have

$$\frac{\partial \phi(\bar{\alpha})}{\partial \alpha_m} = \frac{\partial f(x^*(\bar{\alpha}), \bar{\alpha})}{\partial \alpha_m} - \lambda \frac{\partial g(x^*(\bar{\alpha}), \bar{\alpha})}{\partial \alpha_m}$$

for $m = 1, \ldots, M$, or in matrix notation,

$$\nabla_\alpha \phi(\bar{\alpha}) = \nabla_\alpha f(x^*(\bar{\alpha}), \bar{\alpha}) - \lambda \nabla_\alpha g(x^*(\bar{\alpha}), \bar{\alpha}).$$

See Section M.L of the Mathematical Appendix for a further discussion of this result.[16]

Because prices are parameters in the EMP that enter only the objective function $p \cdot x$, the change in the value function of the EMP with respect to a price change at $\bar{p}$, $\nabla_p e(\bar{p}, u)$, is just the vector of partial derivatives with respect to p of the objective function evaluated at the optimizing vector, $h(\bar{p}, u)$. Hence $\nabla_p e(p, u) = h(p, u)$. ■

The idea behind all three proofs is the same: If we are at an optimum in the EMP, the changes in demand caused by price changes have no first-order effect on the consumer's expenditure. This can be most clearly seen in Proof 2; condition (3.G.2) uses the chain rule to break the total effect of the price change into two effects: a direct effect on expenditure from the change in prices holding demand fixed (the first term) and an indirect effect on expenditure caused by the induced change in demand holding prices fixed (the second term). However, because we are at an expenditure minimizing bundle, the first-order conditions for the EMP imply that this latter effect is zero.

Proposition 3.G.2 summarizes several properties of the price derivatives of the Hicksian demand function $D_p h(p, u)$ that are implied by Proposition 3.G.1 [properties (i) to (iii)]. It also records one additional fact about these derivatives [property (iv)].

Proposition 3.G.2: Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated and strictly convex preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. Suppose also that $h(\cdot, u)$ is continuously differentiable at (p, u), and denote its $L \times L$ derivative matrix by $D_p h(p, u)$. Then

(i) $D_p h(p, u) = D_p^2 e(p, u)$.
(ii) $D_p h(p, u)$ is a negative semidefinite matrix.
(iii) $D_p h(p, u)$ is a symmetric matrix.
(iv) $D_p h(p, u) p = 0$.

Proof: Property (i) follows immediately from Proposition 3.G.1 by differentiation. Properties (ii) and (iii) follow from property (i) and the fact that since $e(p, u)$ is a

16. Proof 2 is essentially a proof of the envelope theorem for the special case where the parameters being changed (in this case, prices) affect only the objective function of the problem.

twice continuously differentiable concave function, it has a symmetric and negative semidefinite Hessian (i.e., second derivative) matrix (see Section M.C of the Mathematical Appendix). Finally, for property (iv), note that because $h(p, u)$ is homogeneous of degree zero in p, $h(\alpha p, u) - h(p, u) = 0$ for all α; differentiating this expression with respect to α yields $D_p h(p, u)p = 0$. [Note that because $h(p, u)$ is homogeneous of degree zero, $D_p h(p, u)p = 0$ also follows directly from Euler's formula; see Section M.B of the Mathematical Appendix.] ■

The negative semidefiniteness of $D_p h(p, u)$ is the differential analog of the compensated law of demand, condition (3.E.5). In particular, the differential version of (3.E.5) is $dp \cdot dh(p, u) \leq 0$. Since $dh(p, u) = D_p h(p, u)\, dp$, substituting gives $dp \cdot D_p h(p, u)\, dp \leq 0$ for all dp; therefore, $D_p h(p, u)$ is negative semidefinite. Note that negative semidefiniteness implies that $\partial h_\ell(p, u)/\partial p_\ell \leq 0$ for all ℓ; that is, compensated own-price effects are nonpositive, a conclusion that we have also derived directly from condition (3.E.5).

The symmetry of $D_p h(p, u)$ is an unexpected property. It implies that compensated price cross-derivatives between any two goods ℓ and k must satisfy $\partial h_\ell(p, u)/\partial p_k = \partial h_k(p, u)/\partial p_\ell$. Symmetry is not easy to interpret in plain economic terms. As emphasized by Samuelson (1947), it is a property just beyond what one would derive without the help of mathematics. Once we know that $D_p h(p, u) = \nabla_p^2 e(p, u)$, the symmetry property reflects the fact that the cross derivatives of a (twice continuously differentiable) function are equal. In intuitive terms, this says that when you climb a mountain, you will cover the same net height regardless of the route.[17] As we discuss in Sections 13.H and 13.J, this path-independence feature is closely linked to the transitivity, or "no-cycling", aspect of rational preferences.

We define two goods ℓ and k to be *substitutes* at (p, u) if $\partial h_\ell(p, u)/\partial p_k \geq 0$ and *complements* if this derivative is nonpositive [when Walrasian demands have these relationships at (p, w), the goods are referred to as *gross* substitutes and *gross* complements at (p, w), respectively]. Because $\partial h_\ell(p, u)/\partial p_\ell \leq 0$, property (iv) of Proposition 3.G.2 implies that there must be a good k for which $\partial h_\ell(p, u)/\partial p_k \geq 0$. Hence, Proposition 3.G.2 implies that every good has at least one substitute.

---

17. To see why this is so, consider the twice continuously differentiable function $f(x, y)$. We can express the change in this function's value from (x', y') to (x'', y'') as the summation (technically, the integral) of two different paths of incremental change: $f(x'', y'') - f(x', y') = \int_{y'}^{y''} [\partial f(x', t)/\partial y]\, dt + \int_{x'}^{x''} [\partial f(s, y'')/\partial x]\, ds$ and $f(x'', y'') - f(x', y') = \int_{x'}^{x''} [\partial f(s, y')/\partial x]\, ds + \int_{y'}^{y''} [\partial f(x'', t)/\partial y]\, dt$. For these two to be equal (as they must be), we should have

$$\int_{y'}^{y''} \left[\frac{\partial f(x'', t)}{\partial y} - \frac{\partial f(x'', t)}{\partial y}\right] dt = \int_{x'}^{x''} \left[\frac{\partial f(s, y'')}{\partial x} - \frac{\partial f(s, y')}{\partial x}\right] ds$$

or

$$\int_{y'}^{y''} \left\{\int_{x'}^{x''} \left[\frac{\partial^2 f(s, t)}{\partial y\, \partial x}\right] ds\right\} dt = \int_{x'}^{x''} \left\{\int_{y'}^{y''} \left[\frac{\partial^2 f(s, t)}{\partial x\, \partial y}\right] dt\right\} ds.$$

So equality of cross-derivatives implies that these two different ways of "climbing the function" yield the same result. Likewise, if the cross-partials were not equal to (x'', y''), then for (x', y') close enough to (x'', y''), the last equality would be violated.

The Hicksian and Walrasian Demand Functions

Although the Hicksian demand function is not directly observable (it has the consumer's utility level as an argument), we now show that $D_p h(p, u)$ can nevertheless be computed from the observable Walrasian demand function $x(p, w)$ (its arguments are all observable in principle). This important result, known as the *Slutsky equation*, means that the properties listed in Proposition 3.G.2 translate into restrictions on the observable Walrasian demand function $x(p, w)$.

Proposition 3.G.3: (*The Slutsky Equation*) Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated and strictly convex preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. Then for all (p, w), and $u = v(p, w)$, we have

$$\frac{\partial h_\ell(p, u)}{\partial p_k} = \frac{\partial x_\ell(p, w)}{\partial p_k} + \frac{\partial x_\ell(p, w)}{\partial w} x_k(p, w) \quad \text{for all } \ell, k \tag{3.G.3}$$

or equivalently, in matrix notation,

$$D_p h(p, u) = D_p x(p, w) + D_w x(p, w) x(p, w)^{\mathrm{T}}. \tag{3.G.4}$$

Proof: Consider a consumer facing the price–wealth pair $(\bar{p}, \bar{w})$ and attaining utility level $\bar{u}$. Note that her wealth level $\bar{w}$ must satisfy $\bar{w} = e(\bar{p}, \bar{u})$. From condition (3.E.4), we know that for all (p, u), $h_\ell(p, u) = x_\ell(p, e(p, u))$. Differentiating this expression with respect to p_k and evaluating it at $(\bar{p}, \bar{u})$, we get

$$\frac{\partial h_\ell(\bar{p}, \bar{u})}{\partial p_k} = \frac{\partial x_\ell(\bar{p}, e(\bar{p}, \bar{u}))}{\partial p_k} + \frac{\partial x_\ell(\bar{p}, e(\bar{p}, \bar{u}))}{\partial w} \frac{\partial e(\bar{p}, \bar{u})}{\partial p_k}.$$

Using Proposition 3.G.1, this yields

$$\frac{\partial h_\ell(\bar{p}, \bar{u})}{\partial p_k} = \frac{\partial x_\ell(\bar{p}, e(\bar{p}, \bar{u}))}{\partial p_k} + \frac{\partial x_\ell(\bar{p}, e(\bar{p}, \bar{u}))}{\partial w} h_k(\bar{p}, \bar{u}).$$

Finally, since $\bar{w} = e(\bar{p}, \bar{u})$ and $h_k(\bar{p}, \bar{u}) = x_k(\bar{p}, e(\bar{p}, \bar{u})) = x_k(\bar{p}, \bar{w})$, we have

$$\frac{\partial h_\ell(\bar{p}, \bar{u})}{\partial p_k} = \frac{\partial x_\ell(\bar{p}, \bar{w})}{\partial p_k} + \frac{\partial x_\ell(\bar{p}, \bar{w})}{\partial w} x_k(\bar{p}, \bar{w}). \blacksquare$$

Figure 3.G.1(a) depicts the Walrasian and Hicksian demand curves for good ℓ as a function of p_ℓ, holding other prices fixed at $\bar{p}_{-\ell}$ [we use $\bar{p}_{-\ell}$ to denote a vector

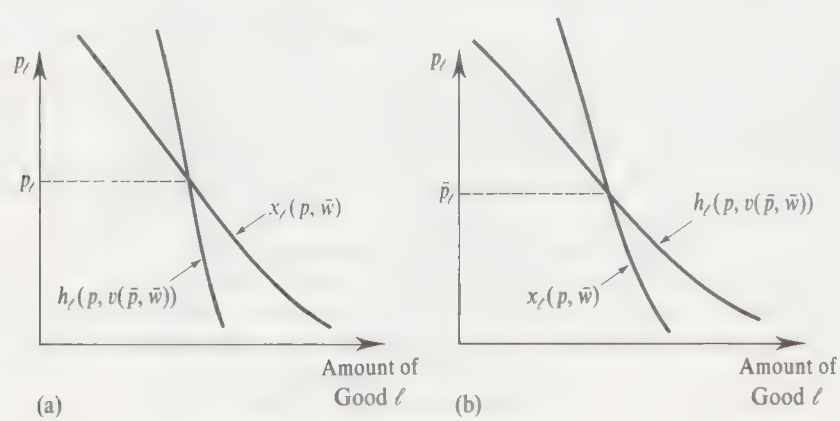


Figure 3.G.1 The Walrasian and Hicksian demand functions for good ℓ. (a) Normal good. (b) Inferior good.

including all prices other than p_ℓ and abuse notation by writing the price vector as $p = (p_\ell, \bar{p}_{-\ell})$]. The figure shows the Walrasian demand function $x(p, \bar{w})$ and the Hicksian demand function $h(p, \bar{u})$ with required utility level $\bar{u} = v((\bar{p}_\ell, \bar{p}_{-\ell}), \bar{w})$. Note that the two demand functions are equal when $p_\ell = \bar{p}_\ell$. The Slutsky equation describes the relationship between the slopes of these two functions at price $\bar{p}_\ell$. In Figure 3.G.1(a), the slope of the Walrasian demand curve at $\bar{p}_\ell$ is less negative than the slope of the Hicksian demand curve at that price. From inspection of the Slutsky equation, this corresponds to a situation where good ℓ is a normal good at $(\bar{p}, \bar{w})$. When p_ℓ increases above $\bar{p}_\ell$, we must increase the consumer's wealth if we are to keep her at the same level of utility. Therefore, if good ℓ is normal, its demand falls by more in the absence of this compensation. Figure 3.G.1(b) illustrates a case in which good ℓ is an inferior good. In this case, the Walrasian demand curve has a more negative slope than the Hicksian curve.

Proposition 3.G.3 implies that the matrix of price derivatives $D_p h(p, u)$ of the Hicksian demand function is equal to the matrix

$$S(p, w) = \begin{bmatrix} s_{11}(p, w) & \cdots & s_{1L}(p, w) \\ \vdots & \ddots & \vdots \\ s_{L1}(p, w) & \cdots & s_{LL}(p, w) \end{bmatrix},$$

with $s_{\ell k}(p, w) = \partial x_\ell(p, w)/\partial p_k + [\partial x_\ell(p, w)/\partial w] x_k(p, w)$. This matrix is known as the *Slutsky substitution matrix*. Note, in particular, that $S(p, w)$ is directly computable from knowledge of the (observable) Walrasian demand function $x(p, w)$. Because $S(p, w) = D_p h(p, u)$, Proposition 3.G.2 implies that when demand is generated from preference maximization, $S(p, w)$ must possess the following three properties: it must be *negative semidefinite, symmetric, and satisfy* $S(p, w)p = 0$.

In Section 2.F, the Slutsky substitution matrix $S(p, w)$ was shown to be the matrix of compensated demand derivatives arising from a different form of wealth compensation, the so-called *Slutsky wealth compensation*. Instead of varying wealth to keep utility fixed, as we do here, Slutsky compensation adjusts wealth so that the initial consumption bundle $\bar{x}$ is just affordable at the new prices. Thus, we have the remarkable conclusion that the *derivative of the Hicksian demand function is equal to the derivative of this alternative Slutsky compensated demand.*

We can understand this result as follows: Suppose we have a utility function $u(\cdot)$ and are at initial position $(\bar{p}, \bar{w})$ with $\bar{x} = x(\bar{p}, \bar{w})$ and $\bar{u} = u(\bar{x})$. As we change prices to p', we want to change wealth in order to compensate for the wealth effect arising from this price change. In principle, the compensation can be done in two ways. By changing wealth by amount $\Delta w_{\text{Slutsky}} = p' \cdot x(\bar{p}, \bar{w}) - \bar{w}$, we leave the consumer just able to afford her initial bundle $\bar{x}$. Alternatively, we can change wealth by amount $\Delta w_{\text{Hicks}} = e(p', \bar{u}) - \bar{w}$ to keep her utility level unchanged. We have $\Delta w_{\text{Hicks}} \le \Delta w_{\text{Slutsky}}$, and the inequality will, in general, be strict for any discrete change (see Figure 3.G.2). But because $\nabla_p e(\bar{p}, \bar{u}) = h(\bar{p}, \bar{u}) = x(\bar{p}, \bar{w})$, these two compensations are *identical* for a differential price change starting at $\bar{p}$. Intuitively, this is due to the same fact that led to Proposition 3.G.1: For a differential change in prices, the total effect on the expenditure required to achieve utility level $\bar{u}$ (the Hicksian compensation level) is simply the direct effect of the price change, assuming that the consumption bundle $\bar{x}$ does not change. But this is precisely the calculation done for Slutsky compensation. Hence, the derivatives of the compensated demand functions that arise from these two compensation mechanisms are the same.

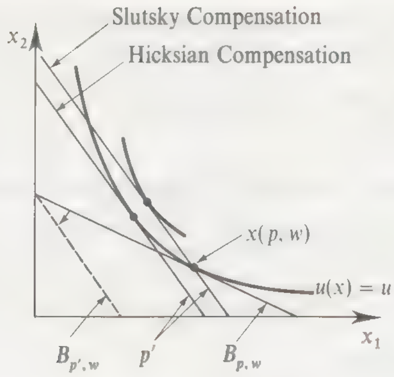


Figure 3.G.2
Hicksian versus Slutsky wealth compensation.

The fact that $D_p h(p, u) = S(p, w)$ allows us to compare the implications of the preference-based approach to consumer demand with those derived in Section 2.F using a choice-based approach built on the weak axiom. Our discussion in Section 2.F concluded that if $x(p, w)$ satisfies the weak axiom (plus homogeneity of degree zero and Walras' law), then $S(p, w)$ is negative semidefinite with $S(p, w)p = 0$. Moreover, we argued that except when $L = 2$, demand satisfying the weak axiom need not have a symmetric Slutsky substitution matrix. Therefore, the results here tell us that the restrictions imposed on demand in the preference-based approach are stronger than those arising in the choice-based theory built on the weak axiom. In fact, it is impossible to find preferences that rationalize demand when the substitution matrix is not symmetric. In Section 3.I, we explore further the role that this symmetry property plays in the relation between the preference and choice-based approaches to demand.

Walrasian Demand and the Indirect Utility Function

We have seen that the minimizing vector of the EMP, $h(p, u)$, is the derivative with respect to p of the EMP's value function $e(p, u)$. The exactly analogous statement for the UMP does not hold. The Walrasian demand, an ordinal concept, cannot equal the price derivative of the indirect utility function, which is not invariant to increasing transformations of utility. But with a small correction in which we normalize the derivatives of $v(p, w)$ with respect to p by the marginal utility of wealth, it holds true. This proposition, called *Roy's identity* (after René Roy), is the parallel result to Proposition 3.G.1 for the demand and value functions of the UMP. As with Proposition 3.G.1, we offer several proofs.

Proposition 3.G.4: (*Roy's Identity*). Suppose that $u(\cdot)$ is a continuous utility function representing a locally nonsatiated and strictly convex preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$. Suppose also that the indirect utility function is differentiable at $(\bar{p}, \bar{w}) \gg 0$. Then

$$x(\bar{p}, \bar{w}) = -\frac{1}{\nabla_w v(\bar{p}, \bar{w})} \nabla_p v(\bar{p}, \bar{w}).$$

That is, for every $\ell = 1, \ldots, L$:

$$x_\ell(\bar{p}, \bar{w}) = -\frac{\partial v(\bar{p}, \bar{w})/\partial p_\ell}{\partial v(\bar{p}, \bar{w})/\partial w}.$$

Proof 1: Let $\bar{u} = v(\bar{p}, \bar{w})$. Because the identity $v(p, e(p, \bar{u})) = \bar{u}$ holds for all p, differentiating with respect to p and evaluating at $p = \bar{p}$ yields

$$\nabla_p v(\bar{p}, e(\bar{p}, \bar{u})) + \frac{\partial v(\bar{p}, e(\bar{p}, \bar{u}))}{\partial w} \nabla_p e(\bar{p}, \bar{u}) = 0.$$

But $\nabla_p e(\bar{p}, \bar{u}) = h(\bar{p}, \bar{u})$ by Proposition 3.G.1, and and so we can substitute and get

$$\nabla_p v(\bar{p}, e(\bar{p}, \bar{u})) + \frac{\partial v(\bar{p}, e(\bar{p}, \bar{u}))}{\partial w} h(\bar{p}, \bar{u}) = 0.$$

Finally, since $\bar{w} = e(\bar{p}, \bar{u})$, we can write

$$\nabla_p v(\bar{p}, \bar{w}) + \frac{\partial v(\bar{p}, \bar{w})}{\partial w} x(\bar{p}, \bar{w}) = 0.$$

Rearranging, this yields the result. ■

Proof 1 of Roy's identity derives the result using Proposition 3.G.1. Proofs 2 and 3 highlight the fact that both results actually follow from the same idea: Because we are at an optimum, the demand response to a price change can be ignored in calculating the effect of a differential price change on the value function. Thus, Roy's identity and Proposition 3.G.1 should be viewed as parallel results for the UMP and EMP. (Indeed, Exercise 3.G.1 asks you to derive Proposition 3.G.1 as a consequence of Roy's identity, thereby showing that the direction of the argument in Proof 1 can be reversed.)

Proof 2: (*First-Order Conditions Argument*). Assume that $x(p, w)$ is differentiable and $x(\bar{p}, \bar{w}) \gg 0$. By the chain rule, we can write

$$\frac{\partial v(\bar{p}, \bar{w})}{\partial p_\ell} = \sum_{k=1}^{L} \frac{\partial u(x(\bar{p}, \bar{w}))}{\partial x_k} \frac{\partial x_k(\bar{p}, \bar{w})}{\partial p_\ell}.$$

Substituting for $\partial u(x(\bar{p}, \bar{w}))/\partial x_k$ using the first-order conditions for the UMP, we have

$$\begin{aligned}\frac{\partial v(\bar{p}, \bar{w})}{\partial p_\ell} &= \sum_{k=1}^{L} \lambda p_k \frac{\partial x_k(\bar{p}, \bar{w})}{\partial p_\ell} \\ &= -\lambda x_\ell(\bar{p}, \bar{w}),\end{aligned}$$

since $\sum_k p_k(\partial x_k(\bar{p}, \bar{w})/\partial p_\ell) = -x_\ell(\bar{p}, \bar{w})$ (Proposition 2.E.2). Finally, we have already argued that $\lambda = \partial v(\bar{p}, \bar{w})/\partial w$ (see Section 3.D); use of this fact yields the result. ■

Proof 2 is again essentially a proof of the envelope theorem, this time for the case where the parameter that varies enters only the constraint. The next result uses the envelope theorem directly.

Proof 3: (*Envelope Theorem Argument*) Applied to the UMP, the envelope theorem tells us directly that the utility effect of a marginal change in p_ℓ is equal to its effect on the consumer's budget constraint weighted by the Lagrange multiplier λ of the consumer's wealth constraint. That is, $\partial v(\bar{p}, \bar{w})/\partial p_\ell = -\lambda x_\ell(\bar{p}, \bar{w})$. Similarly, the utility effect of a differential change in wealth $\partial v(p, w)/\partial w$ is just λ. Combining these two facts yields the result. ■

Proposition 3.G.4 provides a substantial payoff. Walrasian demand is much easier to compute from indirect than from direct utility. To derive $x(p, w)$ from the indirect

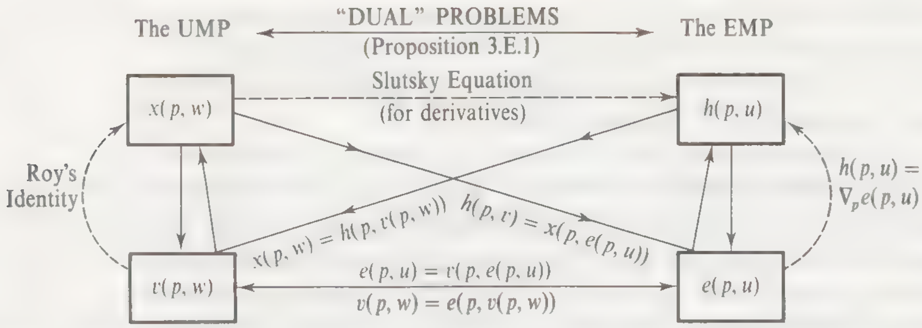


Figure 3.G.3
Relationships between the UMP and the EMP.

utility function, no more than the calculation of derivatives is involved; no system of first-order condition equations needs to be solved. Thus, it may often be more convenient to express tastes in indirect utility form. In Chapter 4, for example, we will be interested in preferences with the property that wealth expansion paths are linear over some range of wealth. It is simple to verify using Roy's identity that indirect utilities of the *Gorman* form $v(p, w) = a(p) + b(p)w$ have this property (see Exercise 3.G.11).

Figure 3.G.3 summarizes the connection between the demand and value functions arising from the UMP and the EMP; a similar figure appears in Deaton and Muellbauer (1980). The solid arrows indicate the derivations discussed in Sections 3.D and 3.E. Starting from a given utility function in the UMP or the EMP, we can derive the optimal consumption bundles $x(p, w)$ and $h(p, u)$ and the value functions $v(p, w)$ and $e(p, u)$. In addition, we can go back and forth between the value functions and demand functions of the two problems using relationships (3.E.1) and (3.E.4).

The relationships developed in this section are represented in Figure 3.G.3 by dashed arrows. We have seen here that the demand vector for each problem can be calculated from its value function and that the derivatives of the Hicksian demand function can be calculated from the observable Walrasian demand using Slutsky's equation.

3.H Integrability

If a continuously differentiable demand function $x(p, w)$ is generated by rational preferences, then we have seen that it must be homogeneous of degree zero, satisfy Walras' law, and have a substitution matrix $S(p, w)$ that is symmetric and negative semidefinite (n.s.d.) at all (p, w). We now pose the reverse question: *If we observe a demand function $x(p, w)$ that has these properties, can we find preferences that rationalize $x(\cdot)$?* As we show in this section (albeit somewhat unrigorously), the answer is yes; these conditions are sufficient for the existence of rational generating preferences. This problem, known as the *integrability problem*, has a long tradition in economic theory, beginning with Antonelli (1886); we follow the approach of Hurwicz and Uzawa (1971).

There are several theoretical and practical reasons why this question and result are of interest.

On a theoretical level, the result tells us two things. First, it tells us that not only are the properties of homogeneity of degree zero, satisfaction of Walras' law, and a

symmetric and negative semidefinite substitution matrix necessary consequences of the preference-based demand theory, but these are also *all* of its consequences. As long as consumer demand satisfies these properties, there is *some* rational preference relation that could have generated this demand.

Second, the result completes our study of the relation between the preference-based theory of demand and the choice-based theory of demand built on the weak axiom. We have already seen, in Section 2.F, that although a rational preference relation always generates demand possessing a symmetric substitution matrix, the weak axiom need not do so. Therefore, we already know that when $S(p, w)$ is not symmetric, demand satisfying the weak axiom cannot be rationalized by preferences. The result studied here tightens this relationship by showing that demand satisfying the weak axiom (plus homogeneity of degree zero and Walras' law) can be rationalized by preferences *if and only if* it has a symmetric substitution matrix $S(p, w)$. Hence, the *only* thing added to the properties of demand by the rational preference hypothesis, beyond what is implied by the weak axiom, homogeneity of degree zero, and Walras' law, is symmetry of the substitution matrix.

On a practical level, the result is of interest for at least two reasons. First, as we shall discuss in Section 3.J, to draw conclusions about welfare effects we need to know the consumer's preferences (or, at the least, her expenditure function). The result tells how and when we can recover this information from observation of the consumer's demand behavior.

Second, when conducting empirical analyses of demand, we often wish to estimate demand functions of a relatively simple form. If we want to allow only functions that can be tied back to an underlying preference relation, there are two ways to do this. One is to specify various utility functions and derive the demand functions that they lead to until we find one that seems statistically tractable. However, the result studied here gives us an easier way; it allows us instead to begin by specifying a tractable demand function and then simply check whether it satisfies the necessary and sufficient conditions that we identify in this section. We do not need to actually derive the utility function; the result allows us to check whether it is, in principle, possible to do so.

The problem of recovering preferences $\succsim$ from $x(p, w)$ can be subdivided into two parts: (i) recovering an expenditure function $e(p, u)$ from $x(p, w)$, and (ii) recovering preferences from the expenditure function $e(p, u)$. Because it is the more straightforward of the two tasks, we discuss (ii) first.

Recovering Preferences from the Expenditure Function

Suppose that $e(p, u)$ is the consumer's expenditure function. By Proposition 3.E.2, it is strictly increasing in u and is continuous, nondecreasing, homogeneous of degree one, and concave in p. In addition, because we are assuming that demand is single-valued, we know that $e(p, u)$ must be differentiable (by Propositions 3.F.1 and 3.G.1).

Given this function $e(p, u)$, how can we recover a preference relation that generates it? Doing so requires finding, for each utility level u, an at-least-as-good-as set $V_u \subset \mathbb{R}^L$ such that $e(p, u)$ is the minimal expenditure required for the consumer to purchase a bundle in V_u at prices $p \gg 0$. That is, we want to identify a set V_u such that, for all

$p \gg 0$, we have

$$e(p, u) = \underset{x \geq 0}{\text{Min}} \quad p \cdot x$$
$$\text{s.t. } x \in V_u.$$

In the framework of Section 3.F, V_u is a set whose support function is precisely $e(p, u)$.

The result in Proposition 3.H.1 shows that the set $V_u = \{x \in \mathbb{R}^L_+ : p \cdot x \geq e(p, u)$ for all $p \gg 0\}$ accomplishes this objective.

Proposition 3.H.1: Suppose that $e(p, u)$ is strictly increasing in u and is continuous, increasing, homogeneous of degree one, concave, and differentiable in p. Then, for every utility level u, $e(p, u)$ is the expenditure function associated with the at-least-as-good-as set

$$V_u = \{x \in \mathbb{R}^L_+ : p \cdot x \geq e(p, u) \text{ for all } p \gg 0\}.$$

That is, $e(p, u) = \text{Min}\,\{p \cdot x : x \in V_u\}$ for all $p \gg 0$.

Proof: The properties of $e(p, u)$ and the definition of V_u imply that V_u is nonempty, closed, and bounded below. Given $p \gg 0$, it can be shown that these conditions insure that $\text{Min}\,\{p \cdot x : x \in V_u\}$. exists. It is immediate from the definition of V_u that $e(p, u) \leq \text{Min}\,\{p \cdot x : x \in V_u\}$. What remains in order to establish the result is to show equality. We do this by showing that $e(p, u) \geq \text{Min}\,\{p \cdot x : x \in V_u\}$.

For any p and p', the concavity of $e(p, u)$ in p implies that (see Section M.C of the Mathematical Appendix)

$$e(p', u) \leq e(p, u) + \nabla_p e(p, u) \cdot (p' - p).$$

Because $e(p, u)$ is homogeneous of degree one in p, Euler's formula tells us that $e(p, u) = p \cdot \nabla_p e(p, u)$. Thus, $e(p', u) \leq p' \cdot \nabla_p e(p, u)$ for all p'. But since $\nabla_p e(p, u) \geq 0$, this means that $\nabla_p e(p, u) \in V_u$. It follows that $\text{Min}\,\{p \cdot x : x \in V_u\} \leq p \cdot \nabla_p e(p, u) = e(p, u)$, as we wanted (the last equality uses Euler's formula once more). This establishes the result. ∎

Given Proposition 3.H.1, we can construct a set V_u for each level of u. Because $e(p, u)$ is strictly increasing in u, it follows that if $u' > u$, then V_u strictly contains $V_{u'}$. In addition, as noted in the proof of Proposition 3.H.1, each V_u is closed, convex, and bounded below. These various at-least-as-good-as sets then define a preference relation $\succsim$ that has $e(p, u)$ as its expenditure function (see Figure 3.H.1).

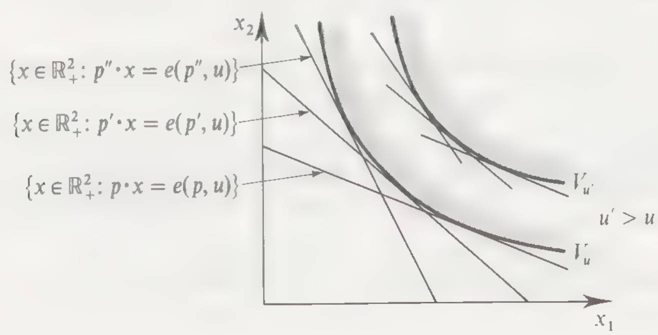


Figure 3.H.1
Recovering preferences from the expenditure function.

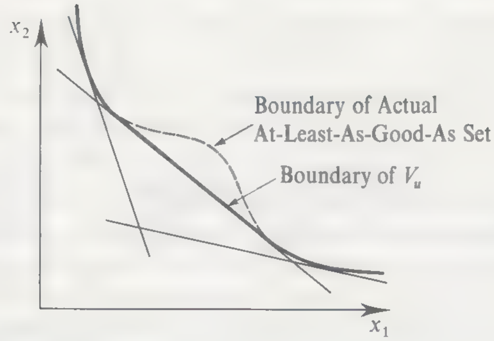


Figure 3.H.2
Recovering [illegible] from the ex[illegible] function when [illegible] consumers' [illegible] are nonconvex.

Proposition 3.H.1 remains valid, with substantially the same proof, when $e(p, u)$ is not differentiable in p. The preference relation constructed as in the proof of the proposition provides a convex preference relation that generates $e(p, u)$. However, it could happen that there are also nonconvex preferences that generate $e(p, u)$. Figure 3.H.2 illustrates a case where the consumer's actual at-least-as-good-as set is nonconvex. The boundary of this set is depicted with a dashed curve. The solid curve shows the boundary of the set $V_u = \{x \in \mathbb{R}^L_+ : p \cdot x \geq e(p, u)$ for all $p \gg 0\}$. Formally, this set is the convex hull of the consumer's actual at-least-as-good-as set, and it also generates the expenditure function $e(p, u)$.

If $e(p, u)$ is differentiable, then any preference relation that generates $e(p, u)$ must be convex. If it were not, then there would be some utility level u and price vector $p \gg 0$ with several expenditure minimizers (see Figure 3.H.2). At this price–utility pair, the expenditure function would not be differentiable in p.

Recovering the Expenditure Function from Demand

It remains to recover $e(p, u)$ from observable consumer behavior summarized in the Walrasian demand $x(p, w)$. We now discuss how this task (which is, more properly, the actual "integrability problem") can be done. We assume throughout that $x(p, w)$ satisfies Walras' law and homogeneity of degree zero and that it is single-valued.

Let us first consider the case of two commodities ($L = 2$). We normalize $p_2 = 1$. Pick an arbitrary price–wealth point $(p_1^0, 1, w^0)$ and assign a utility value of u^0 to bundle $x(p_1^0, 1, w^0)$. We will now recover the value of the expenditure function $e(p_1, 1, u^0)$ at all prices $p_1 > 0$. Because compensated demand is the derivative of the expenditure function with respect to prices (Proposition 3.G.1), recovering $e(\cdot)$ is equivalent to being able to solve (to "integrate") a differential equation with the independent variable p_1 and the dependent variable e. Writing $e(p_1) = e(p_1, 1, u^0)$ and $x_1(p_1, w) = x_1(p_1, 1, w)$ for simplicity, we need to solve the differential equation,

$$\frac{de(p_1)}{dp_1} = x_1(p_1, e(p_1)), \tag{3.H.1}$$

with the initial condition[18] $e(p_1^0) = w^0$.

If $e(p_1)$ solves (3.H.1) for $e(p_1^0) = w^0$, then $e(p_1)$ is the expenditure function associated with the level of utility u^0. Note, in particular, that if the substitution

18. Technically, (3.H.1) is a nonautonomous system in the (p_1, e) plane. Note that p_1 plays the role of the "t" variable.

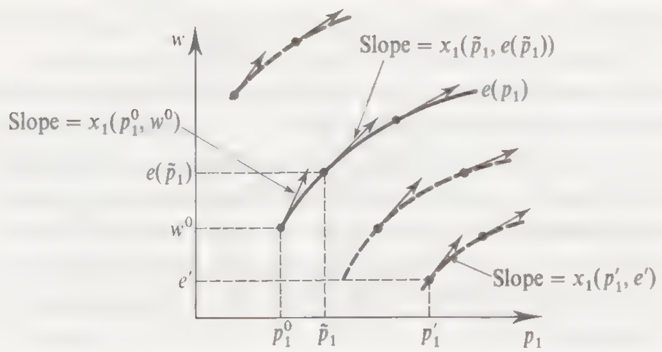


Figure 3.H.3
Recovering the expenditure functions from $x(p, w)$.

matrix is negative semidefinite then $e(p_1)$ will have all the properties of an expenditure function (with the price of good 2 normalized to equal 1). First, because it is the solution to a differential equation, it is by construction continuous in p_1. Second, since $x_1(p, w) \geq 0$, equation (3.H.1) implies that $e(p_1)$ is nondecreasing in p_1. Third, differentiating equation (3.H.1) tells us that

$$\begin{aligned}\frac{d^2 e(p_1)}{dp_1^2} &= \frac{\partial x_1(p_1, 1, e(p_1))}{\partial p_1} + \frac{\partial x_1(p_1, 1, e(p_1))}{\partial w} x_1(p_1, 1, e(p_1)) \\ &= s_{11}(p_1, 1, e(p_1)) \leq 0,\end{aligned}$$

so that the solution $e(p_1)$ is concave in p_1.

Solving equation (3.H.1) is a straightforward problem in ordinary differential equations that, nonetheless, we will not go into. A few weak regularity assumptions guarantee that a solution to (3.H.1) exists for any initial condition (p_1^0, w^0). Figure 3.H.3 describes the essence of what is involved: At each price level p_1 and expenditure level e, we are given a direction of movement with slope $x_1(p_1, e)$. For the initial condition (p_1^0, w^0), the graph of $e(p_1)$ is the curve that starts at (p_1^0, w^0) and follows the prescribed directions of movement.

For the general case of L commodities, the situation becomes more complicated. The (ordinary) differential equation (3.H.1) must be replaced by the system of partial differential equations:

$$\begin{aligned}\frac{\partial e(p)}{\partial p_1} &= x_1(p, e(p)) \\ &\vdots \\ \frac{\partial e(p)}{\partial p_L} &= x_L(p, e(p))\end{aligned} \tag{3.H.2}$$

for initial conditions p^0 and $e(p^0) = w^0$. The existence of a solution to (3.H.2) is *not* automatically guaranteed when $L > 2$. Indeed, if there is a solution $e(p)$, then its Hessian matrix $D_p^2 e(p)$ must be symmetric because the Hessian matrix of any twice continuously differentiable function is symmetric. Differentiating equations (3.H.2), which can be written as $\nabla_p e(p) = x(p, e(p))$, tells us that

$$\begin{aligned}D_p^2 e(p) &= D_p x(p, e(p)) + D_w x(p, e(p)) x(p, e(p))^{\mathrm{T}} \\ &= S(p, e(p)).\end{aligned}$$

Therefore, a necessary condition for the existence of a solution is the symmetry of the Slutsky matrix of $x(p, w)$. This is a comforting fact because we know from previous sections that if market demand is generated from preferences, then the Slutsky matrix is indeed symmetric. It turns out that symmetry of $S(p, w)$ is also sufficient for recovery of the consumer's expenditure function. A basic result of the theory of partial differential equations (called *Frobenius' theorem*) tells us that the symmetry of the $L \times L$ derivative matrix of (3.H.2) at all points of its domain is the necessary and sufficient condition for the existence of a solution to (3.H.2). In addition, if a solution $e(p_1, u_0)$ does exist, then, as long as $S(p, w)$ is negative semidefinite, it will possess the properties of an expenditure function.

We therefore conclude that *the necessary and sufficient condition for the recovery of an underlying expenditure function is the symmetry and negative semidefiniteness of the Slutsky matrix.*[19] Recall from Section 2.F that a differentiable demand function satisfying the weak axiom, homogeneity of degree zero, and Walras' law necessarily has a negative semidefinite Slutsky matrix. Moreover, when $L = 2$, the Slutsky matrix is necessarily symmetric (recall Exercise 2.F.12). Thus, for the case where $L = 2$, we can always find preferences that rationalize any differentiable demand function satisfying these three properties. When $L > 2$, however, the Slutsky matrix of a demand function satisfying the weak axiom (along with homogeneity of degree zero and Walras' law) need not be symmetric; preferences that rationalize a demand function satisfying the weak axiom exist only when it is.

Observe that once we know that $S(p, w)$ is symmetric at all (p, w), we can in fact use (3.H.1) to solve (3.H.2). Suppose that with initial conditions p^0 and $e(p^0) = w^0$, we want to recover $e(\bar{p})$. By changing prices one at a time, we can decompose this problem into L subproblems where only one price changes at each step. Say it is price ℓ. Then with p_k fixed for $k \neq \ell$, the ℓth equation of (3.H.2) is an equation of the form (3.H.1), with the subscript 1 replaced by ℓ. It can be solved by the methods appropriate to (3.H.1). Iterating for different goods, we eventually get to $e(\bar{p})$. It is worthwhile to point out that this method makes mechanical sense even if $S(p, w)$ is not symmetric. However, if $S(p, w)$ is not symmetric (and therefore *cannot* be associated with an underlying preference relation and expenditure function), then the value of $e(\bar{p})$ will *depend on the particular path followed from* p^0 *to* $\bar{p}$ (i.e., on which price is raised first). By this absurdity, the mathematics manage to keep us honest!

3.I Welfare Evaluation of Economic Changes

Up to this point, we have studied the preference-based theory of consumer demand from a positive (behavioral) perspective. In this section, we investigate the normative side of consumer theory, called *welfare analysis.* Welfare analysis concerns itself with the evaluation of the effects of changes in the consumer's environment on her well-being.

Although many of the positive results in consumer theory could also be deduced using an approach based on the weak axiom (as we did in Section 2.F), the preference-based approach to consumer demand is of critical importance for welfare

19. This is subject to minor technical requirements.

analysis. Without it, we would have no means of evaluating the consumer's level of well-being.

In this section, we consider a consumer with a rational, continuous, and locally nonsatiated preference relation $\succsim$. We assume, whenever convenient, that the consumer's expenditure and indirect utility functions are differentiable.

We focus here on the welfare effect of a price change. This is only an example, albeit a historically important one, in a broad range of possible welfare questions one might want to address. We assume that the consumer has a fixed wealth level $w > 0$ and that the price vector is initially p^0. We wish to evaluate the impact on the consumer's welfare of a change from p^0 to a new price vector p^1. For example, some government policy that is under consideration, such as a tax, might result in this change in market prices.[20]

Suppose, to start, that we know the consumer's preferences $\succsim$. For example, we may have derived $\succsim$ from knowledge of her (observable) Walrasian demand function $x(p, w)$, as discussed in Section 3.H. If so, it is a simple matter to determine whether the price change makes the consumer better or worse off: if $v(p, w)$ is any indirect utility function derived from $\succsim$, the consumer is worse off if and only if $v(p^1, w) - v(p^0, w) < 0$.

Although any indirect utility function derived from $\succsim$ suffices for making this comparison, one class of indirect utility functions deserves special mention because it leads to measurement of the welfare change expressed in dollar units. These are called *money metric* indirect utility functions and are constructed by means of the expenditure function. In particular, starting from any indirect utility function $v(\cdot, \cdot)$, choose an arbitrary price vector $\bar{p} \gg 0$, and consider the function $e(\bar{p}, v(p, w))$. This function gives the wealth required to reach the utility level $v(p, w)$ when prices are $\bar{p}$. Note that this expenditure is strictly increasing as a function of the level $v(p, w)$, as shown in Figure 3.I.1. Thus, viewed as a function of (p, w), $e(\bar{p}, v(p, w))$ is itself an indirect utility function for $\succsim$, and

$$e(\bar{p}, v(p^1, w)) - e(\bar{p}, v(p^0, w))$$

provides a measure of the welfare change expressed in dollars.[21]

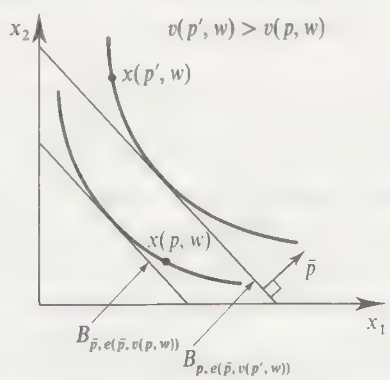


Figure 3.I.1
A money metric indirect utility function.

20. For the sake of expositional simplicity, we do not consider changes that affect wealth here. However, the analysis readily extends to that case (see Exercise 3.I.12).

21. Note that this measure is unaffected by the choice of the initial indirect utility function $v(p, w)$; it depends only on the consumer's preferences $\succsim$ (see Figure 3.I.1).

A money metric indirect utility function can be constructed in this manner for any price vector $\bar{p} \gg 0$. Two particularly natural choices for the price vector $\bar{p}$ are the initial price vector p^0 and the new price vector p^1. These choices lead to two well-known measures of welfare change originating in Hicks (1939), the *equivalent variation* (EV) and the *compensating variation* (CV). Formally, letting $u^0 = v(p^0, w)$ and $u^1 = v(p^1, w)$, and noting that $e(p^0, u^0) = e(p^1, u^1) = w$, we define

$$EV(p^0, p^1, w) = e(p^0, u^1) - e(p^0, u^0) = e(p^0, u^1) - w \tag{3.I.1}$$

and

$$CV(p^0, p^1, w) = e(p^1, u^1) - e(p^1, u^0) = w - e(p^1, u^0). \tag{3.I.2}$$

The equivalent variation can be thought of as the dollar amount that the consumer would be indifferent about accepting in lieu of the price change; that is, it is the change in her wealth that would be *equivalent* to the price change in terms of its welfare impact (so it is negative if the price change would make the consumer worse off). In particular, note that $e(p^0, u^1)$ is the wealth level at which the consumer achieves exactly utility level u^1, the level generated by the price change, at prices p^0. Hence, $e(p^0, u^1) - w$ is the net change in wealth that causes the consumer to get utility level u^1 at prices p^0. We can also express the equivalent variation using the indirect utility function $v(\cdot, \cdot)$ in the following way: $v(p^0, w + EV) = u^1$.[22]

The compensating variation, on the other hand, measures the net revenue of a planner who must *compensate* the consumer for the price change after it occurs, bringing her back to her original utility level u^0. (Hence, the compensating variation is negative if the planner would have to pay the consumer a positive level of compensation because the price change makes her worse off.) It can be thought of as the negative of the amount that the consumer would be just willing to accept from the planner to allow the price change to happen. The compensating variation can also be expressed in the following way: $v(p^1, w - CV) = u^0$.

Figure 3.I.2 depicts the equivalent and compensating variation measures of welfare change. Because both the EV and the CV correspond to measurements of the changes in a money metric indirect utility function, both provide a correct welfare ranking of the alternatives p^0 and p^1; that is, the consumer is better off under p^1 if and only if these measures are positive. In general, however, the specific dollar

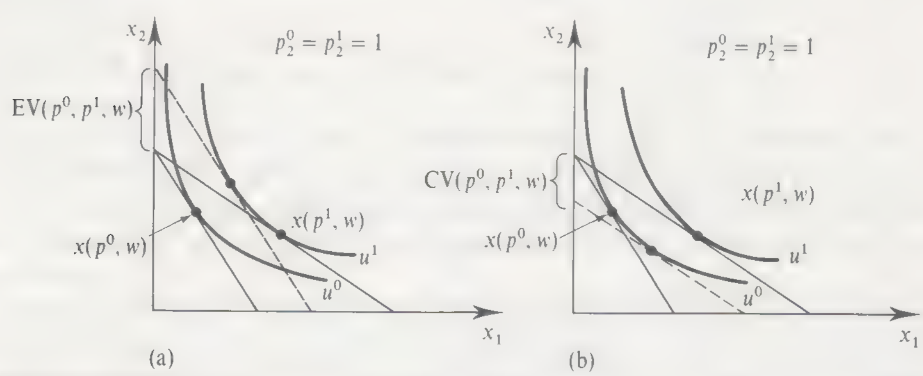


Figure 3.I.2 The equivalent (a) and compensating (b) variation measures of welfare change.

22. Note that if $u^1 = v(p^0, w + EV)$, then $e(p^0, u^1) = e(p^0, v(p^0, w + EV)) = w + EV$. This leads to (3.I.1).

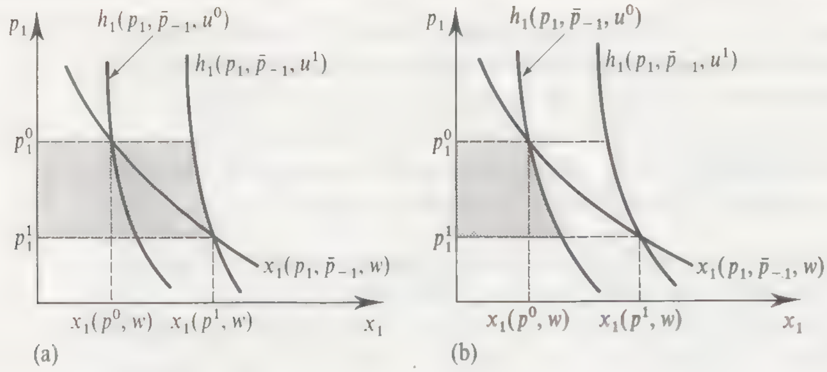


Figure 3.I.3
(a) The equivalent variation.
(b) The compensating variation.

amounts calculated using the EV and CV measures will differ because of the differing price vectors at which compensation is assumed to occur in these two measures of welfare change.

The equivalent and compensating variations have interesting representations in terms of the Hicksian demand curve. Suppose, for simplicity, that only the price of good 1 changes, so that $p_1^0 \neq p_1^1$ and $p_\ell^0 = p_\ell^1 = \bar{p}_\ell$ for all $\ell \neq 1$. Because $w = e(p^0, u^0) = e(p^1, u^1)$ and $h_1(p, u) = \partial e(p, u)/\partial p_1$, we can write

$$\begin{aligned} EV(p^0, p^1, w) &= e(p^0, u^1) - w \\ &= e(p^0, u^1) - e(p^1, u^1) \\ &= \int_{p_1^1}^{p_1^0} h_1(p_1, \bar{p}_{-1}, u^1)\, dp_1, \end{aligned} \tag{3.I.3}$$

where $\bar{p}_{-1} = (\bar{p}_2, \ldots, \bar{p}_L)$. Thus, the change in consumer welfare as measured by the equivalent variation can be represented by the area lying between p_1^0 and p_1^1 and to the left of the Hicksian demand curve for good 1 associated with utility level u^1 (it is equal to this area if $p_1^1 < p_1^0$ and is equal to its negative if $p_1^1 > p_1^0$). The area is depicted as the shaded region in Figure 3.I.3(a).

Similarly, the compensating variation can be written as

$$CV(p^0, p^1, w) = \int_{p_1^1}^{p_1^0} h_1(p_1, \bar{p}_{-1}, u^0)\, dp_1. \tag{3.I.4}$$

Note that we now use the initial utility level u^0. See Figures 3.I.3(b) for its graphic representation.

Figure 3.I.3 depicts a case where good 1 is a normal good. As can be seen in the figure, when this is so, we have $EV(p^0, p^1, w) > CV(p^0, p^1, w)$ (you should check that the same is true when $p_1^1 > p_1^0$). This relation between the EV and the CV reverses when good 1 is inferior (see Exercise 3.I.3). However, if there is no wealth effect for good 1 (e.g., if the underlying preferences are quasilinear with respect to some good $\ell \neq 1$), the CV and EV measures are the *same* because we then have

$$h_1(p_1, \bar{p}_{-1}, u^0) = x_1(p_1, \bar{p}_{-1}, w) = h_1(p_1, \bar{p}_{-1}, u^1).$$

In this case of no wealth effects, we call the common value of CV and EV, which is also the value of the area lying between p_1^0 and p_1^1 and to the left of the market (i.e., Walrasian) demand curve for good 1, the change in *Marshallian consumer surplus*.[23]

23. The term originates from Marshall (1920), who used the area to the left of the market demand curve as a welfare measure in the special case where wealth effects are absent.

Exercise 3.I.1: Suppose that the change from price vector p^0 to price vector p^1 involves a change in the prices of both good 1 (from p_1^0 to p_1^1) and good 2 (from p_2^0 to p_2^1). Express the equivalent variation in terms of the sum of integrals under appropriate Hicksian demand curves for goods 1 and 2. Do the same for the compensating variation measure. Show also that if there are no wealth effects for either good, the compensating and equivalent variations are equal.

Example 3.I.1: *The Deadweight Loss from Commodity Taxation.* Consider a situation where the new price vector p^1 arises because the government puts a tax on some commodity. To be specific, suppose that the government taxes commodity 1, setting a tax on the consumer's purchases of good 1 of t per unit. This tax changes the effective price of good 1 to $p_1^1 = p_1^0 + t$ while prices for all other commodities $\ell \neq 1$ remain fixed at p_ℓ^0 (so we have $p_\ell^1 = p_\ell^0$ for all $\ell \neq 1$). The total revenue raised by the tax is therefore $T = tx_1(p^1, w)$.

An alternative to this commodity tax that raises the same amount of revenue for the government without changing prices is imposition of a "lump-sum" tax of T directly on the consumer's wealth. Is the consumer better or worse off facing this lump-sum wealth tax rather than the commodity tax? She is worse off under the commodity tax if the equivalent variation of the commodity tax $EV(p^0, p^1, w)$, which is negative, is less than $-T$, the amount of wealth she will lose under the lump-sum tax. Put in terms of the expenditure function, this says that she is worse off under commodity taxation if $w - T > e(p^0, u^1)$, so that her wealth after the lump-sum tax is greater than the wealth level that is required at prices p^0 to generate the utility level that she gets under the commodity tax, u^1. The difference $(-T) - EV(p^0, p^1, w) = w - T - e(p^0, u^1)$ is known as the *deadweight loss of commodity taxation.* It measures the extra amount by which the consumer is made worse off by commodity taxation above what is necessary to raise the same revenue through a lump-sum tax.

The deadweight loss measure can be represented in terms of the Hicksian demand curve at utility level u^1. Since $T = tx_1(p^1, w) = th_1(p^1, u^1)$, we can write the deadweight loss as follows [we again let $\bar{p}_{-1} = (\bar{p}_2, \ldots, \bar{p}_L)$, where $p_\ell^0 = p_\ell^1 = \bar{p}_\ell$ for all $\ell \neq 1$]:

$$\begin{aligned}
(-T) - EV(p^0, p^1, w) &= e(p^1, u^1) - e(p^0, u^1) - T \\
&= \int_{p_1^0}^{p_1^0 + t} h_1(p_1, \bar{p}_{-1}, u^1)\, dp_1 - th_1(p_1^0 + t, \bar{p}_{-1}, u^1) \\
&= \int_{p_1^0}^{p_1^0 + t} [h_1(p_1, \bar{p}_{-1}, u^1) - h_1(p_1^0 + t, \bar{p}_{-1}, u^1)]\, dp_1. \qquad (3.I.5)
\end{aligned}$$

Because $h_1(p, u)$ is nonincreasing in p_1, this expression (and therefore the deadweight loss of taxation) is nonnegative, and it is strictly positive if $h_1(p, u)$ is strictly decreasing in p_1. In Figure 3.I.4(a), the deadweight loss is depicted as the area of the crosshatched triangular region. This region is sometimes called the *deadweight loss triangle.*

This deadweight loss measure can also be represented in the commodity space. For example, suppose that $L = 2$, and normalize $p_2^0 = 1$. Consider Figure 3.I.5. Since $(p_1^0 + t)x_1(p^1, w) + p_2^0 x_2(p^1, w) = w$, the bundle $x(p^1, w)$ lies not only on the budget line associated with budget set $B_{p^1, w}$ but also on the budget line associated with budget set $B_{p^0, w-T}$. In contrast, the budget set that generates a utility of u^1 for the consumer at prices p^0 is $B_{p^0, e(p^0, u^1)}$ (or, equivalently,

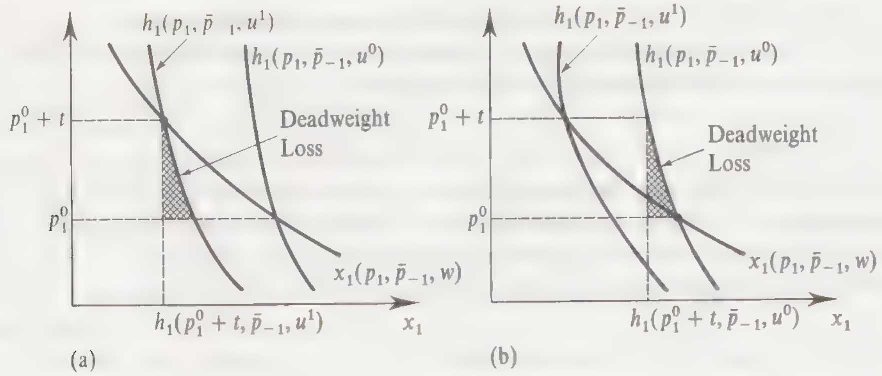


Figure 3.I.4
The deadweight loss from commodity taxation.
(a) Measure based at u^1.
(b) Measure based at u^0.

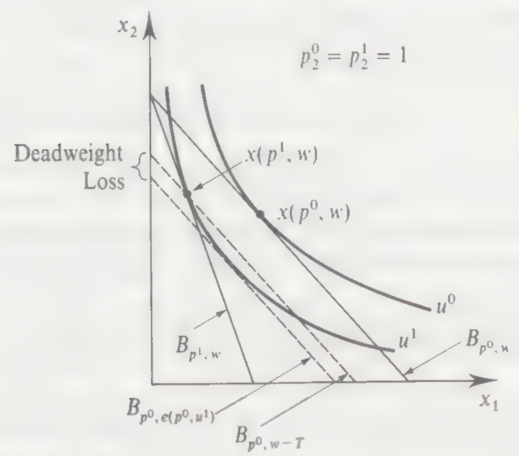


Figure 3.I.5
An alternative depiction of the deadweight loss from commodity taxation.

$B_{p^0, w+EV}$). The deadweight loss is the vertical distance between the budget lines associated with budget sets $B_{p^0, w-T}$ and $B_{p^0, e(p^0, u^1)}$ (recall that $p_2^0 = 1$).

A similar deadweight loss triangle can be calculated using the Hicksian demand curve $h_1(p, u^0)$. It also measures the loss from commodity taxation, but in a different way. In particular, suppose that we examine the surplus or deficit that would arise if the government were to compensate the consumer to keep her welfare under the tax equal to her pretax welfare u^0. The government would run a deficit if the tax collected $th_1(p^1, u^0)$ is less than $-CV(p^0, p^1, w)$ or, equivalently, if $th_1(p^1, u^0) < e(p^1, u^0) - w$. Thus, the deficit can be written as

$$\begin{aligned} -CV(p^0, p^1, w) - th_1(p^1, u) &= e(p^1, u^0) - e(p^0, u^0) - th_1(p^1, u^0) \\ &= \int_{p_1^0}^{p_1^0+t} h_1(p_1, \bar{p}_{-1}, u^0)\, dp_1 - th_1(p_1^0 + t, \bar{p}_{-1}, u^0) \\ &= \int_{p_1^0}^{p_1^0+t} [h_1(p_1, \bar{p}_{-1}, u^0) - h_1(p_1^0 + t, \bar{p}_{-1}, u^0)]\, dp_1 . \end{aligned} \tag{3.I.6}$$

which is again strictly positive as long as $h_1(p, u)$ is strictly decreasing in p_1. This deadweight loss measure is equal to the area of the crosshatched triangular region in Figure 3.I.4(b). ■

Exercise 3.I.2: Calculate the derivative of the deadweight loss measures (3.I.5) and (3.I.6) with respect to t. Show that, evaluated at $t = 0$, these derivatives are equal to zero but that if $h_1(p, u^0)$ is strictly decreasing in p_1, they are strictly positive at all $t > 0$. Interpret.

Up to now, we have considered only the question of whether the consumer was better off at p^1 than at the initial price vector p^0. We saw that both EV and CV provide a correct welfare ranking of p^0 and p^1. Suppose, however, that p^0 is being compared with two possible price vectors p^1 and p^2. In this case, p^1 is better than p^2 if and only if $EV(p^0, p^1, w) > EV(p^0, p^2, w)$, since

$$EV(p^0, p^1, w) - EV(p^0, p^2, w) = e(p^0, u^1) - e(p^0, u^2).$$

Thus, the EV measures $EV(p^0, p^1, w)$ and $EV(p^0, p^2, w)$ can be used not only to compare these two price vectors with p^0 but also to determine which of them is better for the consumer. A comparison of the compensating variations $CV(p^0, p^1, w)$ and $CV(p^0, p^2, w)$, however, will not necessarily rank p^1 and p^2 correctly. The problem is that the CV measure uses the new prices as the base prices in the money metric indirect utility function, using p^1 to calculate $CV(p^0, p^1, w)$ and p^2 to calculate $CV(p^0, p^2, w)$. So

$$CV(p^0, p^1, w) - CV(p^0, p^2, w) = e(p^2, u^0) - e(p^1, u^0),$$

which need not correctly rank p^1 and p^2 [see Exercise 3.I.4 and Chipman and Moore (1980)]. In other words, fixing p^0, $EV(p^0, \cdot, w)$ is a valid indirect utility function (in fact, a money metric one), but $CV(p^0, \cdot, w)$ is not.[24]

An interesting example of the comparison of several possible new price vectors arises when a government is considering which goods to tax. Suppose, for example, that two different taxes are being considered that could raise tax revenue of T: a tax on good 1 of t_1 (creating new price vector p^1) and a tax on good 2 of t_2 (creating new price vector p^2). Note that since they raise the same tax revenue, we have $t_1 x_1(p^1, w) = t_2 x_2(p^2, w) = T$ (see Figure 3.I.6). Because tax t_1

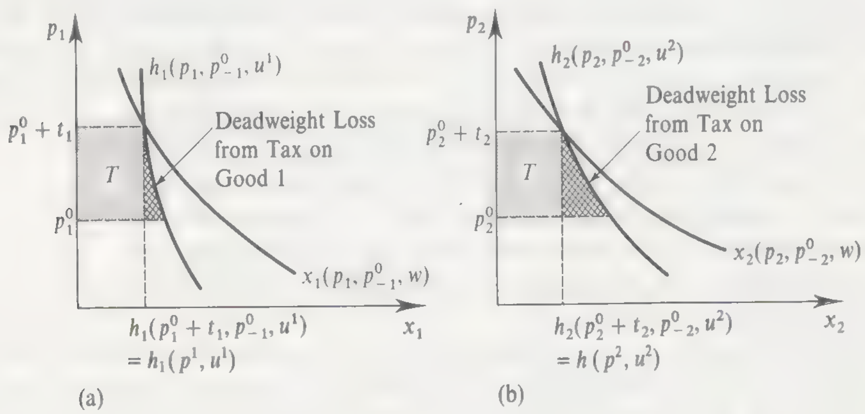


Figure 3.I.6
Comparing two taxes that raise revenue T.
(a) Tax on good 1.
(b) Tax on good 2.

is better than tax t_2 if and only if $EV(p^0, p^1, w) > EV(p^0, p^2, w)$, t_1 is better than t_2 if and only if $[(-T) - EV(p^0, p^1, w)] < [(-T) - EV(p^0, p^2, w)]$, that is, if and only if the deadweight loss arising under tax t_1 is less than that arising under tax t_2.

---

24. Of course, we can rank p^1 and p^2 correctly by seeing whether $CV(p^1, p^2, w)$ is positive or negative.

In summary, if we know the consumer's expenditure function, we can precisely measure the welfare impact of a price change; moreover, we can do it in a convenient way (in dollars). In principle, this might well be the end of the story because, as we saw in Section 3.H, we can recover the consumer's preferences and expenditure function from the observable Walrasian demand function $x(p, w)$.[25] Before concluding, however, we consider two further issues. We first ask whether we may be able to say anything about the welfare effect of a price change when we *do not* have enough information to recover the consumer's expenditure function. We describe a test that provides a sufficient condition for the consumer's welfare to increase from the price change and that uses information only about the two price vectors p^0, p^1 and the initial consumption bundle $x(p^0, w)$. We then conclude by discussing in detail the extent to which the welfare change can be approximated by means of the area to the left of the market (Walrasian) demand curve, a topic of significant historical importance.

Welfare Analysis with Partial Information

In some circumstances, we may not be able to derive the consumer's expenditure function because we may have only limited information about her Walrasian demand function. Here we consider what can be said when the *only* information we possess is knowledge of the two price vectors p^0, p^1 and the consumer's initial consumption bundle $x^0 = x(p^0, w)$. We begin, in Proposition 3.I.1, by developing a simple sufficiency test for whether the consumer's welfare improves as a result of the price change.

Proposition 3.I.1: Suppose that the consumer has a locally nonsatiated rational preference relation $\succsim$. If $(p^1 - p^0)\cdot x^0 < 0$, then the consumer is strictly better off under price–wealth situation (p^1, w) than under (p^0, w).

Proof: The result follows simply from revealed preference. Since $p^0\cdot x^0 = w$ by Walras' law, if $(p^1 - p^0)\cdot x^0 < 0$, then $p^1\cdot x^0 < w$. But if so, x^0 is still affordable under prices p^1 and is, moreover, in the interior of budget set $B_{p^1,w}$. By local nonsatiation, there must therefore be a consumption bundle in $B_{p^1,w}$ that the consumer strictly prefers to x^0. ■

The test in Proposition 3.I.1 can be viewed as a first-order approximation to the true welfare change. To see this, take a first-order Taylor expansion of $e(p, u)$ around the initial prices p^0:

$$e(p^1, u^0) = e(p^0, u^0) + (p^1 - p^0)\cdot\nabla_p e(p^0, u^0) + o(\|p^1 - p^0\|). \qquad (3.I.7)$$

If $(p^1 - p^0)\cdot\nabla_p e(p^0, u^0) < 0$ and the second-order remainder term could be ignored, we would have $e(p^1, u^0) < e(p^0, u^0) = w$, and so we could conclude that the consumer's welfare is greater after the price change. But the concavity of $e(\cdot, u^0)$ in p implies that the remainder term is nonpositive. Therefore, ignoring the remainder term leads to no error here; we do have $e(p^1, u^0) < w$ if $(p^1 - p^0)\cdot\nabla_p e(p^0, u^0) < 0$. Using Proposition 3.G.1 then tells us that $(p^1 - p^0)\cdot\nabla_p e(p^0, u^0) = (p^1 - p^0)\cdot h(p^0, u^0) = (p^1 - p^0)\cdot x^0$, and so we get exactly the test in Proposition 3.I.1.

25. As a practical matter, in applications you should use whatever are the state-of-the-art techniques for performing this recovery.

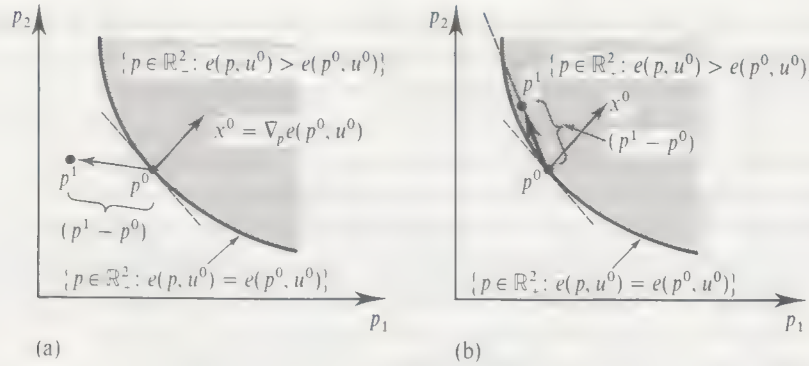


Figure 3.I.7
The welfare ... Propositions ... 3.I.2.
(a) $(p^1 - \ldots$
(b) $(p^1 - \ldots$

What if $(p^1 - p^0)\cdot x^0 > 0$? Can we then say anything about the direction of change in welfare? As a general matter, no. However, examination of the first-order Taylor expansion (3.I.7) tells us that we get a definite conclusion if the price change is, in an appropriate sense, small enough because the remainder term then becomes insignificant relative to the first-order term and can be neglected. This gives the result shown in Proposition 3.I.2.

Proposition 3.I.2: Suppose that the consumer has a differentiable expenditure function. Then if $(p^1 - p^0)\cdot x^0 > 0$, there is a sufficiently small $\bar{\alpha} \in (0, 1)$ such that for all $\alpha < \bar{\alpha}$, we have $e((1 - \alpha)p^0 + \alpha p^1, u^0) > w$, and so the consumer is strictly better off under price–wealth situation (p^0, w) than under $((1 - \alpha)p^0 + \alpha p^1, w)$.

Figure 3.I.7 illustrates these results for the cases where p^1 is such that $(p^1 - p^0)\cdot x^0 < 0$ [panel (a)] and $(p^1 - p^0)\cdot x^0 > 0$ [panel (b)]. In the figure the set of prices $\{p \in \mathbb{R}^2_+ : e(p, u^0) \geq e(p^0, u^0)\}$ is drawn in price space. The concavity of $e(\cdot, u)$ gives it the shape depicted. The initial price vector p^0 lies in this set. By Proposition 3.G.1, the gradient of the expenditure function at this point, $\nabla_p e(p^0, u^0)$, is equal to x^0, the initial consumption bundle. The vector $(p^1 - p^0)$ is the vector connecting point p^0 to the new price point p^1. Figure 3.I.7(a) shows a case where $(p^1 - p^0)\cdot x^0 < 0$. As can be seen there, p^1 lies outside of the set $\{p \in \mathbb{R}^2_+ : e(p, u^0) \geq e(p^0, u^0)\}$, and so we must have $e(p^0, u^0) > e(p^1, u^0)$. In Figure 3.I.7(b), on the other hand, we show a case where $(p^1 - p^0)\cdot x^0 > 0$. Proposition 3.I.2 can be interpreted as asserting that in this case if $(p^1 - p^0)$ is small enough, then $e(p^0, u^0) < e(p^1, u^0)$. This can be seen in Figure 3.I.7(b), because if $(p^1 - p^0)\cdot x^0 > 0$ and p^1 is close enough to p^0 [in the ray with direction $p^1 - p^0$], then price vector p^1 lies in the set $\{p \in \mathbb{R}^2_+ : e(p, u^0) > e(p^0, u^0)\}$.

Using the Area to the Left of the Walrasian (Market) Demand Curve as an Approximate Welfare Measure

Improvements in computational abilities have made the recovery of the consumer's preferences/expenditure function from observed demand behavior, along the lines discussed in Section 3.I, far easier than was previously the case.[26] Traditionally,

26. They have also made it much easier to estimate complicated demand systems that are explicitly derived from utility maximization and from which the parameters of the expenditure function can be derived directly.

however, it has been common practice in applied analyses to rely on approximations of the true welfare change.

We have already seen in (3.I.3) and (3.I.4) that the welfare change induced by a change in the price of good 1 can be exactly computed by using the area to the left of an appropriate Hicksian demand curve. However, these measures present the problem of not being directly observable. A simpler procedure that has seen extensive use appeals to the Walrasian (market) demand curve instead. We call this estimate of welfare change the *area variation* measure (or AV):

$$AV(p^0, p^1, w) = \int_{p_1^1}^{p_1^0} x_1(p_1, \bar{p}_{-1}, w)\, dp_1. \tag{3.I.8}$$

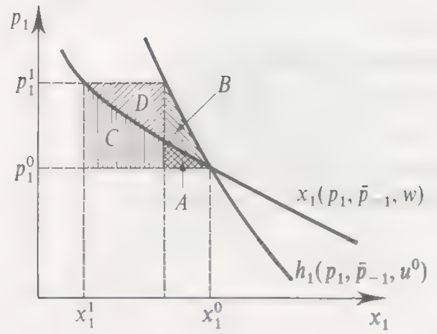
If there are no wealth effects for good 1, then, as we have discussed, $x_1(p, w) = h_1(p, u^0) = h_1(p, u^1)$ for all p and the area variation measure is exactly equal to the equivalent and compensating variation measures. This corresponds to the case studied by Marshall (1920) in which the marginal utility of numeraire is constant. In this circumstance, where the AV measure gives an exact measure of welfare change, the measure is known as the change in *Marshallian consumer surplus.*

More generally, as Figures 3.I.3(a) and 3.I.3(b) make clear, when good 1 is a normal good, the area variation measure overstates the compensating variation and understates the equivalent variation (convince yourself that this is true both when p_1 falls and when p_1 rises). When good 1 is inferior, the reverse relations hold. Thus, when evaluating the welfare change from a change in prices of several goods, or when comparing two different possible price changes, the area variation measure need not give a correct evaluation of welfare change (e.g., see Exercise 3.I.10).

Naturally enough, however, if the wealth effects for the goods under consideration are small, the approximation errors are also small and the area variation measure is almost correct. Marshall argued that if a good is just one commodity among many, then because one extra unit of wealth will spread itself around, the wealth effect for the commodity is bound to be small; therefore, no significant errors will be made by evaluating the welfare effects of price changes for that good using the area measure. This idea can be made precise; for an advanced treatment, see Vives (1987). It is important, however, not to fall into the fallacy of composition; if we deal with a large number of commodities, then while the approximating error may be small for each individually, it may nevertheless not be small in the aggregate.

If $(p_1^1 - p_1^0)$ is small, then the error involved using the area variation measure becomes small as a fraction of the true welfare change. Consider, for example, the compensating variation.[27] In Figure 3.I.8, we see that the area $B + D$, which measures the difference between the area variation and the true compensating variation, becomes small as a fraction of the true compensating variation when $(p_1^1 - p_1^0)$ is small. This might seem to suggest that the area variation measure is a good approximation of the compensating variation measure for small price changes. Note, however, that the same property would hold if instead of the Walrasian demand

27. All the points that follow apply to the equivalent variation as well.



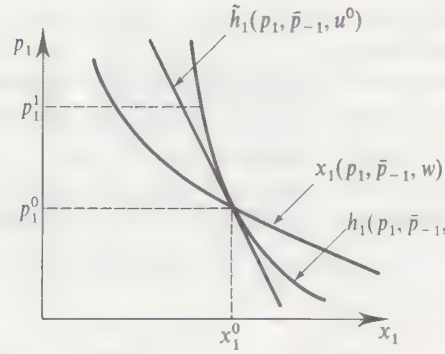


Figure 3.I.8 (lef
The error in
area variation
of welfare chang

Figure 3.I.9 (right
A first-order
approximation of
$h(p, u^0)$ at p^0.

function we were to use *any* function that takes the value $x_1(p_1^0, p_{-1}^0, w)$ at p_1^0.[28] In fact, the approximation error may be quite large *as a fraction of the deadweight loss* [this point is emphasized by Hausman (1981)]. In Figure 3.I.8, for example, the deadweight loss calculated using the Walrasian demand curve is the area $A + C$, whereas the real one is the area $A + B$. The percentage difference between these two areas need not grow small as the price change grows small.[29]

When $(p_1^1 - p_1^0)$ is small, there is a superior approximation procedure available In particular, suppose we take a first-order Taylor approximation of $h(p, u^0)$ at p^0

$$\tilde{h}(p, u^0) = h(p^0, u^0) + D_p h(p^0, u^0)(p - p^0)$$

and we calculate

$$\int_{p_1^1}^{p_1^0} \tilde{h}_1(p_1, \bar{p}_{-1}, u^0)\, dp_1 \tag{3.I.9}$$

as our approximation of the welfare change. The function $\tilde{h}_1(p_1, \bar{p}_{-1}, u^0)$ is depicted in Figure 3.I.9. As can be seen in the figure, because $\tilde{h}_1(p_1, \bar{p}_{-1}, u^0)$ has the same slope as the true Hicksian demand function $h_1(p, u^0)$ at p^0, for small price changes this approximation comes closer than expression (3.I.8) to the true welfare change (and in contrast with the area variation measure, it provides an adequate approximation to the deadweight loss). Because the Hicksian demand curve is the first derivative of the expenditure function, this first-order expansion of the Hicksian demand function at p^0 is, in essence, a second-order expansion of the expenditure function around p^0. Thus, this approximation can be viewed as the natural extension of the first-order test discussed above; see expression (3.I.7).

The approximation in (3.I.9) is directly computable from knowledge of the observable Walrasian demand function $x_1(p, w)$. To see this, note that because $h(p^0, u^0) = x(p^0, w)$ and $D_p h(p^0, u^0) = S(p^0, w)$, $\tilde{h}(p, u^0)$ can be expressed solely in terms that involve the Walrasian demand function and its derivatives at the point

28. In effect, the property identified here amounts to saying that the Walrasian demand function provides a first-order approximation to the compensating variation. Indeed, note that the derivatives of $CV(p^1, p^0, w)$, $EV(p^1, p^0, w)$, and $AV(p^1, p^0, w)$ with respect to p_1^1 evaluated at p_1^0 are all precisely $x_1(p_1^0, p_{-1}^0, w)$.

29. Thus, for example, in the problem discussed above where we compare the deadweight losses induced by taxes on two different commodities that both raise revenue T, the area variation measure need not give the correct ranking even for small taxes.

(p^0, w):

$$\tilde{h}(p, u^0) = x(p^0, w) + S(p^0, w)(p - p^0).$$

In particular, since only the price of good 1 is changing, we have

$$\tilde{h}_1(p_1, \bar{p}_{-1}, u^0) = x_1(p_1^0, \bar{p}_{-1}, w) + s_{11}(p_1^0, \bar{p}_{-1}, w)(p_1 - p_1^0),$$

where

$$s_{11}(p_1^0, \bar{p}_{-1}, w) = \frac{\partial x_1(p^0, w)}{\partial p_1} + \frac{\partial x_1(p^0, w)}{\partial w} x_1(p^0, w).$$

When $(p^1 - p^0)$ is small, this procedure provides a better approximation to the true compensating variation than does the area variation measure. However, if $(p^1 - p^0)$ is large, we cannot tell which is the better approximation. It is entirely possible for the area variation measure to be superior. After all, its use guarantees some sensitivity of the approximation to demand behavior away from p^0, whereas the use of $\tilde{h}(p, u^0)$ does not.

3.J The Strong Axiom of Revealed Preference

We have seen that in the context of consumer demand theory, consumer choice may satisfy the weak axiom but not be capable of being generated by a rational preference relation (see Sections 2.F and 3.G). We could therefore ask: Can we find a necessary and sufficient consistency condition on consumer demand behavior that is in the same style as the WA but that does imply that demand behavior can be rationalized by preferences? The answer is "yes", and it was provided by Houthakker (1950) in the form of the *strong axiom of revealed preference* (*SA*), a kind of recursive closure of the weak axiom.[30]

Definition 3.J.1: The market demand function $x(p, w)$ satisfies the *strong axiom of revealed preference* (the SA) if for any list

$$(p^1, w^1), \ldots, (p^N, w^N)$$

with $x(p^{n+1}, w^{n+1}) \neq x(p^n, w^n)$ for all $n \leq N - 1$, we have $p^N \cdot x(p^1, w^1) > w^N$ whenever $p^n \cdot x(p^{n+1}, w^{n+1}) \leq w^n$ for all $n \leq N - 1$.

In words, if $x(p^1, w^1)$ is *directly or indirectly revealed preferred* to $x(p^N, w^N)$, then $x(p^N, w^N)$ cannot be (directly) revealed preferred to $x(p^1, w^1)$ [so $x(p^1, w^1)$ cannot be affordable at (p^N, w^N)]. For example, the SA was violated in Example 2.F.1. It is clear that the SA is satisfied if demand originates in rational preferences. The converse is a deeper result. It is stated in Proposition 3.J.1; the proof, which is advanced, is presented in small type.

Proposition 3.J.1: If the Walrasian demand function $x(p, w)$ satisfies the strong axiom of revealed preference then there is a rational preference relation $\succsim$ that rationalizes $x(p, w)$, that is, such that for all (p, w), $x(p, w) \succ y$ for every $y \neq x(p, w)$ with $y \in B_{p,w}$.

30. For an informal account of revealed preference theory after Samuelson, see Mas-Colell (1982).

Proof: We follow Richter (1966). His proof is based on set theory and differs markedly from the differential equations techniques used originally by Houthakker.[31]

Define a relation $\succ^1$ on commodity vectors by letting $x \succ^1 y$ whenever $x \neq y$ and we have $x = x(p, w)$ and $p \cdot y \leq w$ for some (p, w). The relation $\succ^1$ can be read as "directly revealed preferred to." From $\succ^1$ define a new relation $\succ^2$, to be read as "directly or indirectly revealed preferred to," by letting $x \succ^2 y$ whenever there is a chain $x^1 \succ^1 x^2 \succ^1, \ldots, \succ^1 x^N$ with $x^1 = x$ and $x^N = y$. Observe that, by construction, $\succ^2$ is transitive. According to the SA, $\succ^2$ is also irreflexive (i.e., $x \succ^2 x$ is impossible). A certain axiom of set theory (known as Zorn's lemma) tells us the following: *Every relation $\succ^2$ that is transitive and irreflexive* (called a *partial order*) *has a total extension* $\succ^3$, an irreflexive and transitive relation such that, first, $x \succ^2 y$ implies $x \succ^3 y$ and, second, whenever $x \neq y$, we have either $x \succ^3 y$ or $y \succ^3 x$. Finally, we can define $\succsim$ by letting $x \succsim y$ whenever $x = y$ or $x \succ^3 y$. It is not difficult now to verify that $\succsim$ is complete and transitive and that $x(p, w) \succ y$ whenever $p \cdot y \leq w$ and $y \neq x(p, w)$. ■

The proof of Proposition 3.J.1 uses only the single-valuedness of $x(p, w)$. Provided choice is single-valued, the same result applies to the abstract theory of choice of Chapter 1. The fact that the budgets are competitive is immaterial.

In Exercise 3.J.1, you are asked to show that the WA is equivalent to the SA when $L = 2$. Hence, by Proposition 3.J.1, when $L = 2$ and demand satisfies the WA, we can always find a rationalizing preference relation, a result that we have already seen in Section 3.H. When $L > 2$, however, the SA is stronger than the WA. In fact, Proposition 3.J.1 tells us that a choice-based theory of demand founded on the strong axiom is essentially equivalent to the preference-based theory of demand presented in this chapter.

The strong axiom is therefore essentially equivalent both to the rational preference hypothesis and to the symmetry and negative semidefiniteness of the Slutsky matrix. We have seen that the weak axiom is essentially equivalent to the negative semidefiniteness of the Slutsky matrix. It is therefore natural to ask whether there is an assumption on preferences that is weaker than rationality and that leads to a theory of consumer demand equivalent to that based on the WA. Violations of the SA mean cycling choice, and violations of the symmetry of the Slutsky matrix generate path dependence in attempts to "integrate back" to preferences. This suggests preferences that may violate the transitivity axiom. See the appendix with W. Shafer in Kihlstrom, Mas-Colell, and Sonnenschein (1976) for further discussion of this point.

APPENDIX A: CONTINUITY AND DIFFERENTIABILITY PROPERTIES OF WALRASIAN DEMAND

In this appendix, we investigate the continuity and differentiability properties of the Walrasian demand correspondence $x(p, w)$. We assume that $x \gg 0$ for all $(p, w) \gg 0$ and $x \in x(p, w)$.

31. Yet a third approach, based on linear programming techniques, was provided by Afriat (1967).

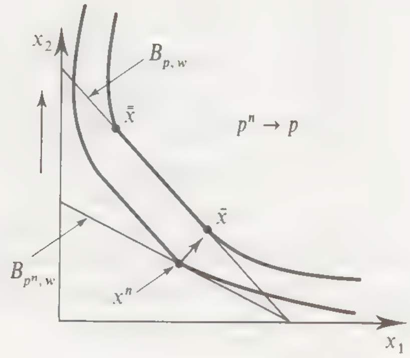


Figure 3.AA.1
An upper hemicontinuous Walrasian demand correspondence.

Continuity

Because $x(p, w)$ is, in general, a correspondence, we begin by introducing a generalization of the more familiar continuity property for functions, called *upper hemicontinuity*.

Definition 3.AA.1: The Walrasian demand correspondence $x(p, w)$ is *upper hemicontinuous* at $(\bar{p}, \bar{w})$ if whenever $(p^n, w^n) \to (\bar{p}, \bar{w})$, $x^n \in x(p^n, w^n)$ for all n, and $x = \lim_{n\to\infty} x^n$, we have $x \in x(p, w)$.[32]

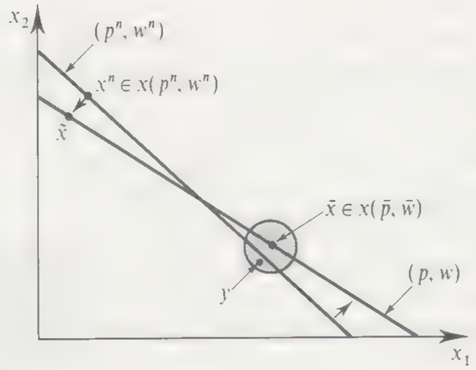
In words, a demand correspondence is upper hemicontinuous at $(\bar{p}, \bar{w})$ if for any sequence of price–wealth pairs the limit of any sequence of optimal demand bundles is optimal (although not necessarily uniquely so) at the limiting price–wealth pair. If $x(p, w)$ is single-valued at all $(p, w) \gg 0$, this notion is equivalent to the usual continuity property for functions.

Figure 3.AA.1 depicts an upper hemicontinuous demand correspondence: When $p^n \to p$, $x(\cdot, w)$ exhibits a jump in demand behavior at the price vector p, being x^n for all p^n but suddenly becoming the interval of consumption bundles $[\bar{x}, \bar{\bar{x}}]$ at p. It is upper hemicontinuous because $\bar{x}$ (the limiting optimum for p^n along the sequence) is an element of segment $[\bar{x}, \bar{\bar{x}}]$ (the set of optima at price vector p). See Section M.H of the Mathematical Appendix for further details on upper hemicontinuity.

Proposition 3.AA.1: Suppose that $u(\cdot)$ is a continuous utility function representing locally nonsatiated preferences $\succsim$ on the consumption set $X = \mathbb{R}^L_+$. Then the derived demand correspondence $x(p, w)$ is upper hemicontinuous at all $(p, w) \gg 0$. Moreover, if $x(p, w)$ is a function [i.e., if $x(p, w)$ has a single element for all (p, w)], then it is continuous at all $(p, w) \gg 0$.

Proof: To verify upper hemicontinuity, suppose that we had a sequence $\{(p^n, w^n)\}_{n=1}^{\infty} \to (\bar{p}, \bar{w}) \gg 0$ and a sequence $\{x^n\}_{n=1}^{\infty}$ with $x^n \in x(p^n, w^n)$ for all n, such that $x^n \to \tilde{x}$ and $\tilde{x} \notin x(\bar{p}, \bar{w})$. Because $p^n \cdot x^n \leq w^n$ for all n, taking limits as $n \to \infty$, we conclude that $\bar{p} \cdot \tilde{x} \leq \bar{w}$. Thus, $\tilde{x}$ is a feasible consumption bundle when the budget set is $B_{\bar{p},\bar{w}}$. However, since it is not optimal in this set, it must be that $u(\bar{x}) > u(\tilde{x})$ for some $\bar{x} \in B_{\bar{p},\bar{w}}$.

32. We use the notation $z^n \to z$ as synonymous with $z = \lim_{n\to\infty} z^n$. This definition of upper hemicontinuity applies only to correspondences that are "locally bounded" (see Section M.H of the Mathematical Appendix). Under our assumptions, the Walrasian demand correspondence satisfies this property at all $(p, w) \gg 0$.



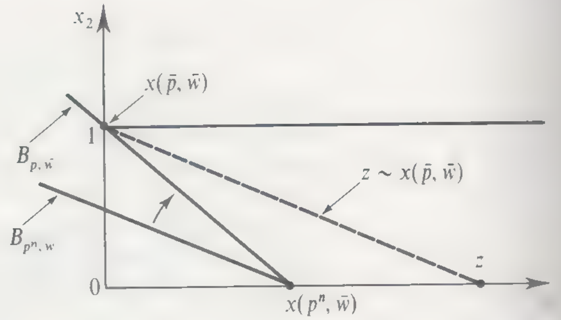


Figure 3.AA.2 (left)
Finding a bundle y such that $p \cdot y < w$ and $u(y) > u(\tilde{x})$.

Figure 3.AA.3 (right)
The locally cheaper test fails at price–wealth pair $(\bar{p}, \bar{w}) = (1, \bar{w}, \bar{w})$.

By the continuity of $u(\cdot)$, there is a y arbitrarily close to $\tilde{x}$ such that $p \cdot y < w$ and $u(y) > u(\tilde{x})$. This bundle y is illustrated in Figure 3.AA.2.

Note that if n is large enough, we will have $p^n \cdot y < w^n$ [since $(p^n, w^n) \to (p, w)$]. Hence, y is an element of the budget set B_{p^n, w^n}, and we must have $u(x^n) \geq u(y)$ because $x^n \in x(p^n, w^n)$. Taking limits as $n \to \infty$, the continuity of $u(\cdot)$ then implies that $u(\tilde{x}) \geq u(y)$, which gives us a contradiction. We must therefore have $\tilde{x} \in x(p, w)$, establishing upper hemicontinuity of $x(p, w)$.

The same argument also establishes continuity if $x(p, w)$ is in fact a function. ■

Suppose that the consumption set is an arbitrary closed set $X \subset \mathbb{R}^L_+$. Then the continuity (or upper hemicontinuity) property still follows at any $(\bar{p}, \bar{w})$ that passes the following (*locally cheaper consumption*) test: "Suppose that $x \in X$ is affordable (i.e., $\bar{p} \cdot x \leq \bar{w}$). Then there is a $y \in X$ arbitrarily close to x and that costs less than $\bar{w}$ (i.e., $\bar{p} \cdot y < \bar{w}$)." For example, in Figure 3.AA.3, commodity 2 is available only in indivisible unit amounts. The locally cheaper test then fails at the price–wealth point $(\bar{p}, \bar{w}) = (1, \bar{w}, \bar{w})$, where a unit of good 2 becomes just affordable. You can easily verify by examining the figure [in which the dashed line indicates indifference between the points (0, 1) and z] that demand will fail to be upper hemicontinuous when $p_2 = \bar{w}$. In particular, for price–wealth points $(p^n, \bar{w})$ such that $p_1^n = 1$ and $p_2^n > \bar{w}$, $x(p^n, \bar{w})$ involves only the consumption of good 1; whereas at $(\bar{p}, \bar{w}) = (1, \bar{w}, \bar{w})$, we have $x(\bar{p}, \bar{w}) = (0, 1)$. Note that the proof of Proposition 3.AA.1 fails when the locally cheaper consumption condition does not hold because we cannot find a consumption bundle y with the properties described there.

Differentiability

Proposition 3.AA.1 has established that if $x(p, w)$ is a function, then it is continuous. Often it is convenient that it be differentiable as well. We now discuss when this is so. We assume for the remaining paragraphs that $u(\cdot)$ is strictly quasiconcave and twice continuously differentiable and that $\nabla u(x) \neq 0$ for all x.

As we have shown in Section 3.D, the first-order conditions for the UMP imply that $x(p, w) \gg 0$ is, for some $\lambda > 0$, the unique solution of the system of $L + 1$ equations in $L + 1$ unknowns:

$$\nabla u(x) - \lambda p = 0$$
$$p \cdot x - w = 0.$$

Therefore, the *implicit function theorem* (see Section M.E of the Mathematical Appendix) tells us that the differentiability of the solution $x(p, w)$ as a function of the parameters (p, w) of the system depends on the Jacobian matrix of this system having a nonzero determinant. The Jacobian matrix [i.e., the derivative matrix of the $L + 1$ component functions with respect to the $L + 1$ variables (x, λ)] is

$$\begin{bmatrix} D^2u(x) & -p \\ p^{\mathrm{T}} & 0 \end{bmatrix}.$$

Since $\nabla u(x) = \lambda p$ and $\lambda > 0$, the determinant of this matrix is nonzero if and only if the determinant of the *bordered Hessian* of $u(x)$ at x is nonzero:

$$\begin{vmatrix} D^2u(x) & \nabla u(x) \\ [\nabla u(x)]^{\mathrm{T}} & 0 \end{vmatrix} \neq 0.$$

This condition has a straightforward geometric interpretation. It means that the indifference set through x has a nonzero curvature at x; it is not (even infinitesimally) flat. This condition is a slight technical strengthening of strict quasiconcavity [just as the strictly concave function $f(x) = -(x^4)$ has $f''(0) = 0$, a strictly quasiconcave function could have a bordered Hessian determinant that is zero at a point].

We conclude, therefore, that $x(p, w)$ is differentiable *if and only if* the determinant of the bordered Hessian of $u(\cdot)$ is nonzero at $x(p, w)$. It is worth noting the following interesting fact (which we shall not prove here): If $x(p, w)$ is differentiable at (p, w), then the Slutsky matrix $S(p, w)$ has maximal possible rank; that is, the rank of $S(p, w)$ equals $L - 1$.[33]

REFERENCES

Afriat, S. (1967). The construction of utility functions from expenditure data. *International Economic Review* **8**: 67–77.

Antonelli, G. B. (1886). *Sulla Teoria Matematica della Economia Politica.* Pisa: Nella tipografia del Folchetto. [English translation: On the mathematical theory of political economy.] In *Preferences, Utility and Demand*, edited by J. Chipman, L. Hurwicz, and H. Sonnenschein. New York: Harcourt Brace Jovanovich, 1971.]

Chipman, J., and J. Moore. (1980). Compensating variation, consumer's surplus, and welfare. *American Economic Review* **70**: 933–48.

Deaton, A., and J. Muellbauer (1980). *Economics and Consumer Behavior.* Cambridge, U.K.: Cambridge University Press.

Debreu, G. (1960). Topological methods in cardinal utility. In *Mathematical Methods in the Social Studies, 1959*, edited by K. Arrow, S. Karlin, and P. Suppes. Stanford, Calif.: Stanford University Press.

Diewert, W. E. (1982). Duality approaches to microeconomic theory. Chap. 12 in *Handbook of Mathematical Economics.* Vol. 2, edited by K. Arrow and M. Intriligator. Amsterdam: North-Holland.

Green, J. R., and W. Heller. (1981). Mathematical analysis and convexity with applications to economics. Chap. 1 in *Handbook of Mathematical Economics*, Vol. 1, edited by K. Arrow and M. Intriligator. Amsterdam. North-Holland.

---

33. This statement applies only to demand generated from a twice continuously differentiable utility function. It need not be true when this condition is not met. For example, the demand function $x(p, w) = (w/(p_1 + p_2), w/(p_1 + p_2))$ is differentiable, and it is generated by the utility function $u(x) = \text{Min}\,\{x_1, x_2\}$, which is not twice continuously differentiable at all x. The substitution matrix for this demand function has all its entries equal to zero and therefore has rank equal to zero.

Hausman, J. (1981). Exact consumer surplus and deadweight loss. *American Economic Review* **71**: 662–76.
Hicks, J. (1939). *Value and Capital*. Oxford: Clarendon Press.
Houthakker, H. S. (1950). Revealed preference and the utility function. *Economica* **17**: 159–74.
Hurwicz, L., and Uzawa. (1971). On the integrability of demand functions. Chap. 6 in *Preferences, Utility and Demand*, edited by J. Chipman, L. Hurwicz, and H. Sonnenschein. New York: Harcourt Brace, Jovanovich.
Kihlstrom, R., A. Mas-Colell, and H. Sonnenschein. (1976). The demand theory of the weak axiom of revealed preference. *Econometrica* **44**: 971–78.
McKenzie, L. (1956–57). Demand theory without a utility index. *Review of Economic Studies* **24**: 185–89.
Marshall, A. (1920). *Principles of Economics*. London: Macmillan.
Mas-Colell, A. (1982). Revealed preference after Samuelson, in *Samuelson and Neoclassical Economics*, edited by G. Feiwel. Boston: Kluwer-Nijhoff.
Richter, M. (1966). Revealed preference theory. *Econometrica* **34**: 635–45.
Samuelson, P. (1947). *Foundations of Economics Analysis*. Cambridge, Mass.: Harvard University Press.
Slutsky, E. (1915). Sulla teoria del bilancio del consumatore. *Giornali degli Economisti* **51**: 1–26. [English translation: On the theory of the budget of the consumer, in *Readings in Price Theory*, edited by G. Stigler and K. Boulding. Chicago: Richard Irwin, 1952.]
Stone, J. E. (1954). Linear expenditure systems and demand analysis: An application to the pattern of British demand. *Economic Journal* **64**: 511–27.
Vives, X. (1987). Small income effects: A Marshallian theory of consumer surplus and downward sloping demand. *Review of Economic Studies* **54**: 87–103.

EXERCISES

3.B.1^{A} In text.

3.B.2^{B} The preference relation $\succsim$ defined on the consumption set $X = \mathbb{R}^L_+$ is said to be *weakly monotone* if and only if $x \geq y$ implies that $x \succsim y$. Show that if $\succsim$ is transitive, locally nonsatiated, and weakly monotone, then it is monotone.

3.B.3^{A} Draw a convex preference relation that is locally nonsatiated but is not monotone.

3.C.1^{B} Verify that the lexicographic ordering is complete, transitive, strongly monotone, and strictly convex.

3.C.2^{B} Show that if $u(\cdot)$ is a continuous utility function representing $\succsim$, then $\succsim$ is continuous.

3.C.3^{C} Show that if for every x the upper and lower contour sets $\{y \in \mathbb{R}^L_+ : y \succsim x\}$ and $\{y \in \mathbb{R}^L_+ : x \succsim y\}$ are closed, then $\succsim$ is continuous according to Definition 3.C.1.

3.C.4^{B} Exhibit an example of a preference relation that is not continuous but is representable by a utility function.

3.C.5^{C} Establish the following two results:

(a) A continuous $\succsim$ is homothetic if and only if it admits a utility function $u(x)$ that is homogeneous of degree one; i.e., $u(\alpha x) = \alpha u(x)$ for all $\alpha > 0$.

(b) A continuous $\succsim$ on $(-\infty, \infty) \times \mathbb{R}^{L-1}_+$ is quasilinear with respect to the first commodity if and only if it admits a utility function $u(x)$ of the form $u(x) = x_1 + \phi(x_2, \ldots, x_L)$. [*Hint*: The existence of some continuous utility representation is guaranteed by Proposition 3.G.1.]

After answering (a) and (b), argue that these properties of $u(\cdot)$ are cardinal.

3.C.6^B Suppose that in a two-commodity world, the consumer's utility function takes the form $u(x) = [\alpha_1 x_1^\rho + \alpha_2 x_2^\rho]^{1/\rho}$. This utility function is known as the *constant elasticity of substitution* (or *CES*) utility function.

(a) Show that when $\rho = 1$, indifference curves become linear.

(b) Show that as $\rho \to 0$, this utility function comes to represent the same preferences as the (generalized) Cobb–Douglas utility function $u(x) = x_1^{\alpha_1} x_2^{\alpha_2}$.

(c) Show that as $\rho \to -\infty$, indifference curves become "right angles"; that is, this utility function has in the limit the indifference map of the Leontief utility function $u(x_1, x_2) = \text{Min}\,\{x_1, x_2\}$.

3.D.1^A In text.

3.D.2^A In text.

3.D.3^B Suppose that $u(x)$ is differentiable and strictly quasiconcave and that the Walrasian demand function $x(p, w)$ is differentiable. Show the following:

(a) If $u(x)$ is homogeneous of degree one, then the Walrasian demand function $x(p, w)$ and the indirect utility function $v(p, w)$ are homogeneous of degree one [and hence can be written in the form $x(p, w) = w\tilde{x}(p)$ and $v(p, w) = w\tilde{v}(p)$] and the wealth expansion path (see Section 2.E) is a straight line through the origin. What does this imply about the wealth elasticities of demand?

(b) If $u(x)$ is strictly quasiconcave and $v(p, w)$ is homogeneous of degee one in w, then $u(x)$ must be homogeneous of degree one.

3.D.4^B Let $(-\infty, \infty) \times \mathbb{R}_+^{L-1}$ denote the consumption set, and assume that preferences are strictly convex and quasilinear. Normalize $p_1 = 1$.

(a) Show that the Walrasian demand functions for goods $2, \ldots, L$ are independent of wealth. What does this imply about the wealth effect (see Section 2.E) of demand for good 1?

(b) Argue that the indirect utility function can be written in the form $v(p, w) = w + \phi(p)$ for some function $\phi(\cdot)$.

(c) Suppose, for simplicity, that $L = 2$, and write the consumer's utility function as $u(x_1, x_2) = x_1 + \eta(x_2)$. Now, however, let the consumption set be $\mathbb{R}_+^2$ so that there is a nonnegativity constraint on consumption of the numeraire x_1. Fix prices p, and examine how the consumer's Walrasian demand changes as wealth w varies. When is the nonnegativity constraint on the numeraire irrelevant?

3.D.5^B Consider again the CES utility function of Exercise 3.C.6, and assume that $\alpha_1 = \alpha_2 = 1$.

(a) Compute the Walrasian demand and indirect utility functions for this utility function.

(b) Verify that these two functions satisfy all the properties of Propositions 3.D.2 and 3.D.3.

(c) Derive the Walrasian demand correspondence and indirect utility function for the case of linear utility and the case of Leontief utility (see Exercise 3.C.6). Show that the CES Walrasian demand and indirect utility functions approach these as ρ approaches 1 and $-\infty$, respectively.

(d) The *elasticity of substitution between goods* 1 and 2 is defined as

$$\xi_{12}(p, w) = -\frac{\partial[x_1(p, w)/x_2(p, w)]}{\partial[p_1/p_2]} \frac{p_1/p_2}{x_1(p, w)/x_2(p, w)}.$$

Show that for the CES utility function, $\xi_{12}(p, w) = 1/(1 - \rho)$, thus justifying its name. What is $\xi_{12}(p, w)$ for the linear, Leontief, and Cobb–Douglas utility functions?

3.D.6^B Consider the three-good setting in which the consumer has utility function $u(x) = (x_1 - b_1)^\alpha (x_2 - b_2)^\beta (x_3 - b_3)^\gamma$.

(a) Why can you assume that $\alpha + \beta + \gamma = 1$ without loss of generality? Do so for the rest of the problem.

(b) Write down the first-order conditions for the UMP, and derive the consumer's Walrasian demand and indirect utility functions. This system of demands is known as the *linear expenditure system* and is due to Stone (1954).

(c) Verify that these demand functions satisfy the properties listed in Propositions 3.D.2 and 3.D.3.

3.D.7^B There are two commodities. We are given two budget sets B_{p^0, w^0} and B_{p^1, w^1} described, respectively, by $p^0 = (1, 1)$, $w^0 = 8$ and $p^1 = (1, 4)$, $w^1 = 26$. The observed choice at (p^0, w^0) is $x^0 = (4, 4)$. At (p^1, w^1), we have a choice x^1 such that $p \cdot x^1 = w^1$.

(a) Determine the region of permissible choices x^1 if the choices x^0 and x^1 are consistent with maximization of preferences.

(b) Determine the region of permissible choices x^1 if the choices x^0 and x^1 are consistent with maximization of preferences that are quasilinear with respect to the *first* good.

(c) Determine the region of permissible choices x^1 if the choices x^0 and x^1 are consistent with maximization of preferences that are quasilinear with respect to the *second* good.

(d) Determine the region of permissible choices x^1 if the choices x^0 and x^1 are consistent with maximization of preferences for which both goods are normal.

(e) Determine the region of permissible choices x^1 if the choices x^0 and x^1 are consistent with maximization of homothetic preferences.

[*Hint*: The ideal way to answer this exercise relies on (good) pictures as much as possible.]

3.D.8^A Show that for all (p, w), $w\, \partial v(p, w)/\partial w = -p \cdot \nabla_p v(p, w)$.

3.E.1^A In text.

3.E.2^A In text.

3.E.3^B Prove that a solution to the EMP exists if $p \gg 0$ and there is some $x \in \mathbb{R}^L_+$ satisfying $u(x) \geq u$.

3.E.4^B Show that if the consumer's preferences $\succsim$ are convex, then $h(p, u)$ is a convex set. Also show that if $u(x)$ is strictly convex, then $h(p, u)$ is single-valued.

3.E.5^B Show that if $u(\cdot)$ is homogeneous of degree one, then $h(p, u)$ and $e(p, u)$ are homogeneous of degree one in u [i.e., they can be written as $h(p, u) = \tilde{h}(p)u$ and $e(p, u) = \tilde{e}(p)u$].

3.E.6^B Consider the constant elasticity of substitution utility function studied in Exercises 3.C.6 and 3.D.5 with $\alpha_1 = \alpha_2 = 1$. Derive its Hicksian demand function and expenditure function. Verify the properties of Propositions 3.E.2 and 3.E.3.

3.E.7^B Show that if $\succsim$ is quasilinear with respect to good 1, the Hicksian demand functions for goods $2, \ldots, L$ do not depend on u. What is the form of the expenditure function in this case?

3.E.8^A For the Cobb–Douglas utility function, verify that the relationships in (3.E.1) and (3.E.4) hold. Note that the expenditure function can be derived by simply inverting the indirect utility function, and vice versa.

3.E.9^B Use the relations in (3.E.1) to show that the properties of the indirect utility function identified in Proposition 3.D.3 imply Proposition 3.E.2. Likewise, use the relations in (3.E.1) to prove that Proposition 3.E.2 implies Proposition 3.D.3.

3.E.10^B Use the relations in (3.E.1) and (3.E.4) and the properties of the indirect utility and expenditure functions to show that Proposition 3.D.2 implies Proposition 3.E.4. Then use these facts to prove that Proposition 3.E.3 implies Proposition 3.D.2.

3.F.1^B Prove formally that a closed, convex set $K \subset \mathbb{R}^L$ equals the intersection of the half-spaces that contain it (use the separating hyperplane theorem).

3.F.2^A Show by means of a graphic example that the separating hyperplane theorem does not hold for nonconvex sets. Then argue that if K is closed and not convex, there is always some $x \notin K$ that cannot be separated from K.

3.G.1^B Prove that Proposition 3.G.1 is implied by Roy's identity (Proposition 3.G.4).

3.G.2^B Verify for the case of a Cobb–Douglas utility function that all of the propositions in Section 3.G hold.

3.G.3^B Consider the (linear expenditure system) utility function given in Exercise 3.D.6.

(a) Derive the Hicksian demand and expenditure functions. Check the properties listed in Propositions 3.E.2 and 3.E.3.

(b) Show that the derivatives of the expenditure function are the Hicksian demand function you derived in (a).

(c) Verify that the Slutsky equation holds.

(d) Verify that the own-substitution terms are negative and that compensated cross-price effects are symmetric.

(e) Show that $S(p, w)$ is negative semidefinite and has rank 2.

3.G.4^B A utility function $u(x)$ is *additively separable* if it has the form $u(x) = \sum_\ell u_\ell(x_\ell)$.

(a) Show that additive separability is a cardinal property that is preserved only under linear transformations of the utility function.

(b) Show that the induced ordering on any group of commodities is independent of whatever fixed values we attach to the remaining ones. It turns out that this ordinal property is not only necessary but also sufficient for the existence of an additive separable representation. [You should *not* attempt a proof. This is very hard. See Debreu (1960)].

(c) Show that the Walrasian and Hicksian demand functions generated by an additively separable utility function admit no inferior goods if the functions $u_\ell(\cdot)$ are strictly concave. (You can assume differentiability and interiority to answer this question.)

(d) (Harder) Suppose that all $u_\ell(\cdot)$ are identical and twice differentiable. Let $\hat{u}(\cdot) = u_\ell(\cdot)$. Show that if $-[t\hat{u}''(t)/\hat{u}'(t)] < 1$ for all t, then the Walrasian demand $x(p, w)$ has the so-called *gross substitute property*, i.e., $\partial x_\ell(p, w)/\partial p_k > 0$ for all ℓ and $k \neq \ell$.

3.G.5^C (*Hicksian composite commodities.*) Suppose there are two groups of desirable commodities, x and y, with corresponding prices p and q. The consumer's utility function is $u(x, y)$, and her wealth is $w > 0$. Suppose that prices for goods y always vary in proportion to one another, so that we can write $q = \alpha q_0$. For any number $z \geq 0$, define the function

$$\tilde{u}(x, z) = \underset{y}{\text{Max}} \quad u(x, y)$$
$$\text{s.t. } q_0 \cdot y \leq z.$$

(a) Show that if we imagine that the goods in the economy are x and a single composite commodity z, that $\tilde{u}(x, z)$ is the consumer's utility function, and that α is the price of the composite commodity, then the solution to $\text{Max}_{x,z}\, \tilde{u}(x, z)$ s.t. $p \cdot x + \alpha z \leq w$ will give the consumer's actual levels of x and $z = q_0 \cdot y$.

(b) Show that properties of Walrasian demand functions identified in Propositions 3.D.2 and 3.G.4 hold for $x(p, \alpha, w)$ and $z(p, \alpha, w)$.

(c) Show that the properties in Propositions 3.E.3, and 3.G.1 to 3.G.3 hold for the Hicksian demand functions derived using $\tilde{u}(x, z)$.

3.G.6[B] (F. M. Fisher) A consumer in a three-good economy (goods denoted x_1, x_2, and x_3; prices denoted p_1, p_2, p_3) with wealth level $w > 0$ has demand functions for commodities 1 and 2 given by

$$x_1 = 100 - 5\frac{p_1}{p_3} + \beta\frac{p_2}{p_3} + \delta\frac{w}{p_3}$$

$$x_2 = \alpha + \beta\frac{p_1}{p_3} + \gamma\frac{p_2}{p_3} + \delta\frac{w}{p_3}$$

where Greek letters are nonzero constants.

(a) Indicate how to calculate the demand for good 3 (but do not actually do it).

(b) Are the demand functions for x_1 and x_2 appropriately homogeneous?

(c) Calculate the restrictions on the numerical values of α, β, γ and δ implied by utility maximization.

(d) Given your results in part (c), for a fixed level of x_3 draw the consumer's indifference curve in the x_1, x_2 plane.

(e) What does your answer to (d) imply about the form of the consumer's utility function $u(x_1, x_2, x_3)$?

3.G.7[A] A striking duality is obtained by using the concept of *indirect demand function.* Fix w at some level, say $w = 1$; from now on, we write $x(p, 1) = x(p)$, $v(p, 1) = v(p)$. The *indirect demand function* $g(x)$ is the inverse of $x(p)$; that is, it is the rule that assigns to every commodity bundle $x \gg 0$ the price vector $g(x)$ such that $x = x(g(x), 1)$. Show that

$$g(x) = \frac{1}{x \cdot \nabla u(x)} \nabla u(x).$$

Deduce from Proposition 3.G.4 that

$$x(p) = \frac{1}{p \cdot \nabla v(p)} \nabla v(p).$$

Note that this is a completely symmetric expression. Thus, direct (Walrasian) demand is the normalized derivative of indirect utility, and indirect demand is the normalized derivative of direct utility.

3.G.8[B] The indirect utility function $v(p, w)$ is logarithmically homogeneous if $v(p, \alpha w) = v(p, w) + \ln \alpha$ for $\alpha > 0$ [in other words, $v(p, w) = \ln(v^*(p, w))$, where $v^*(p, w)$ is homogeneous of degree one]. Show that if $v(\cdot, \cdot)$ is logarithmically homogeneous, then $x(p, 1) = -\nabla_p v(p, 1)$.

3.G.9[C] Compute the Slutsky matrix from the indirect utility function.

3.G.10[B] For a function of the Gorman form $v(p, w) = a(p) + b(p)w$, which properties will the functions $a(\cdot)$ and $b(\cdot)$ have to satisfy for $v(p, w)$ to qualify as an indirect utility function?

3.G.11^B Verify that an indirect utility function in Gorman form exhibits linear wealth-expansion curves.

3.G.12^B What restrictions on the Gorman form correspond to the cases of homothetic and quasilinear preferences?

3.G.13^C Suppose that the indirect utility function $v(p, w)$ is a polynomial of degree n on w (with coefficients that may depend on p). Show that any individual wealth-expansion path is contained in a linear subspace of at most dimension $n + 1$. Interpret.

3.G.14^A The matrix below records the (Walrasian) demand substitution effects for a consumer endowed with rational preferences and consuming three goods at the prices $p_1 = 1$, $p_2 = 2$, and $p_3 = 6$:

$$\begin{bmatrix} -10 & ? & ? \\ ? & -4 & ? \\ 3 & ? & ? \end{bmatrix}.$$

Supply the missing numbers. Does the resulting matrix possess all the properties of a substitution matrix?

3.G.15^B Consider the utility function

$$u = 2x_1^{1/2} + 4x_2^{1/2}.$$

(a) Find the demand functions for goods 1 and 2 as they depend on prices and wealth.

(b) Find the compensated demand function $h(\cdot)$.

(c) Find the expenditure function, and verify that $h(p, u) = \nabla_p e(p, u)$.

(d) Find the indirect utility function, and verify Roy's identity.

3.G.16^C Consider the expenditure function

$$e(p, u) = \exp\left\{\sum_\ell \alpha_\ell \log p_\ell + \left(\prod_\ell p_\ell^{\beta_\ell}\right)u\right\}.$$

(a) What restrictions on $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ are necessary for this to be derivable from utility maximization?

(b) Find the indirect utility that corresponds to it.

(c) Verify Roy's identity and the Slutsky equation.

3.G.17^B [From Hausman (1981)] Suppose $L = 2$. Consider a "local" indirect utility function defined in some neighborhood of price–wealth pair $(\bar{p}, \bar{w})$ by

$$v(p, w) = -\exp(-bp_1/p_2)\left[\frac{w}{p_2} + \frac{1}{b}\left(a\frac{p_1}{p_2} + \frac{a}{b} + c\right)\right].$$

(a) Verify that the local demand function for the first good is

$$x_1(p, w) = a\frac{p_1}{p_2} + b\frac{w}{p_2} + c.$$

(b) Verify that the local expenditure function is

$$e(p, u) = -p_2 u \exp(bp_1/p_2) - \frac{1}{b}\left(ap_1 + \frac{a}{b}p_2 + cp_2\right).$$

(c) Verify that the local Hicksian demand function for the first commodity is

$$h_1(p, u) = -ub \exp(bp_1/p_2) - \frac{a}{b}.$$

3.G.18^C Show that every good is related to every other good by a chain of (weak) substitutes; that is, for any goods ℓ and k, either $\partial h_\ell(p, u)/\partial p_k \geq 0$, or there exists a good r such that $\partial h_\ell(p, u)/\partial p_r \geq 0$ and $\partial h_r(p, u)/\partial p_k \geq 0$, or there is . . . , and so on. [*Hint*: Argue first the case of two commodities. Use next the insights on composite commodities gained in Exercise 3.G.5 to handle the case of three, and then L, commodities.]

3.H.1^C Show that if $e(p, u)$ is continuous, increasing in u, homogeneous of degree one, nondecreasing, and concave in p, then the utility function $u(x) = \text{Sup}\,\{u: x \in V_u\}$, where $V_u = \{y: p \cdot y \geq e(p, u) \text{ for all } p \gg 0\}$, defined for $x \gg 0$, satisfies $e(p, u) = \text{Min}\,\{p \cdot x: u(x) \geq u\}$ for any $p \gg 0$.

3.H.2^B Use Proposition 3.F.1 to argue that if $e(p, u)$ is differentiable in p, then there are no (strongly monotone) nonconvex preferences generating $e(\cdot)$.

3.H.3^A How would you recover $v(p, w)$ from $e(p, u)$?

3.H.4^B Suppose that we are given as primitive, not the Walrasian demand but the indirect demand function $g(x)$ introduced in Exercise 3.G.7. How would you go about recovering $\succsim$? Restrict yourself to the case $L = 2$.

3.H.5^B Suppose you know the indirect utility function. How would you recover from it the expenditure function and the direct utility function?

3.H.6^B Suppose that you observe the Walrasian demand functions $x_\ell(p, w) = \alpha_\ell w/p_\ell$ for all $\ell = 1, \ldots, L$ with $\sum_\ell \alpha_\ell = 1$. Derive the expenditure function of this demand system. What is the consumer's utility function?

3.H.7^B Answer the following questions with reference to the demand function in Exercise 2.F.17.

(a) Let the utility associated with consumption bundle $x = (1, 1, \ldots, 1)$ be 1. What is the expenditure function $e(p, 1)$ associated with utility level $u = 1$? [*Hint*: Use the answer to (d) in Exercise 2.F.17.]

(b) What is the upper contour set of consumption bundle $x = (1, 1, \ldots, 1)$?

3.I.1^B In text.

3.I.2^B In text.

3.I.3^B Consider a price change from initial price vector p^0 to new price vector $p^1 \leq p^0$ in which only the price of good ℓ changes. Show that $CV(p^0, p^1, w) > EV(p^0, p^1, w)$ if good ℓ is inferior.

3.I.4^B Construct an example in which a comparison of $CV(p^0, p^1, w)$ and $CV(p^0, p^2, w)$ does not give the correct welfare ranking of p^1 versus p^2.

3.I.5^B Show that if $u(x)$ is quasilinear with respect to the first good (and we fix $p_1 = 1$), then $CV(p^0, p^1, w) = EV(p^0, p^1, w)$ for any (p^0, p^1, w).

3.I.6^A Suppose there are $i = 1, \ldots, I$ consumers with utility functions $u_i(x)$ and wealth w_i. We consider a change from p^0 to p^1. Show that if $\sum_i CV_i(p^0, p^1, w_i) > 0$ then we can find $\{w_i'\}_{i=1}^I$ such that $\sum_i w_i' \leq \sum_i w_i$ and $v_i(p^1, w_i') \geq v_i(p^0, w_i)$ for all i. That is, it is in principle possible to compensate everybody for the change in prices.

3.I.7[B] There are three commodities (i.e., $L = 3$), of which the third is a numeraire (let $p_3 = 1$). The market demand function $x(p, w)$ has

$$x_1(p, w) = a + bp_1 + cp_2$$
$$x_2(p, w) = d + ep_1 + gp_2.$$

(a) Give the parameter restrictions implied by utility maximization.

(b) Estimate the equivalent variation for a change of prices from $(p_1, p_2) = (1, 1)$ to $(\bar{p}_1, \bar{p}_2) = (2, 2)$. Verify that without appropriate symmetry, there is no path independence. Assume symmetry for the rest of the exercise.

(c) Let EV_1, EV_2, and EV be the equivalent variations for a change of prices from $(p_1, p_2) = (1, 1)$ to, respectively, $(2, 1)$, $(1, 2)$, and $(2, 2)$. Compare EV with $EV_1 + EV_2$ as a function of the parameters of the problem. Interpret.

(d) Suppose that the price increases in (c) are due to taxes. Denote the deadweight losses for each of the three experiments by DW_1, DW_2, and DW. Compare DW with $DW_1 + DW_2$ as a function of the parameters of the problem.

(e) Suppose the initial tax situation has prices $(p_1, p_2) = (1, 1)$. The government wants to raise a fixed (small) amount of revenue R through commodity taxes. Call t_1 and t_2 the tax rates for the two commodities. Determine the optimal tax rates as a function of the parameters of demand if the optimality criterion is the minimization of deadweight loss.

3.I.8[B] Suppose we are in a three-commodity market (i.e. $L = 3$). Letting $p_3 = 1$, the demand functions for goods 1 and 2 are

$$x_1(p, w) = a_1 + b_1 p_1 + c_1 p_2 + d_1 p_1 p_2$$
$$x_2(p, w) = a_2 + b_2 p_1 + c_2 p_2 + d_2 p_1 p_2.$$

(a) Note that the demand for goods 1 and 2 does not depend on wealth. Write down the most general class of utility functions whose demand has this property.

(b) Argue that if the demand functions in (a) are generated from utility maximization, then the values of the parameters cannot be arbitrary. Write down as exhaustive a list as you can of the restrictions implied by utility maximization. Justify your answer.

(c) Suppose that the conditions in (b) hold. The initial price situation is $p = (p_1, p_2)$, and we consider a change to $p' = (p'_1, p'_2)$. Derive a measure of welfare change generated in going from p to p'.

(d) Let the values of the parameters be $a_1 = a_2 = 3/2$, $b_1 = c_2 = 1$, $c_1 = b_2 = 1/2$, and $d_1 = d_2 = 0$. Suppose the initial price situation is $p = (1, 1)$. Compute the equivalent variation for a move to p' for each of the following three cases: (i) $p' = (2, 1)$, (ii) $p' = (1, 2)$, and (iii) $p' = (2, 2)$. Denote the respective answers by EV_1, EV_2, EV_3. Under which condition will you have $EV_3 = EV_1 + EV_2$? Discuss.

3.I.9[B] In a one-consumer economy, the government is considering putting a tax of t per unit on good ℓ and rebating the proceeds to the consumer (who nonetheless does not consider the effect of her purchases on the size of the rebate). Suppose that $s_{\ell\ell}(p, w) < 0$ for all (p, w). Show that the optimal tax (in the sense of maximizing the consumer's utility) is zero.

3.I.10[B] Construct an example in which the area variation measure approach incorrectly ranks p^0 and p^1. [*Hint*: Let the change from p^0 to p^1 involve a change in the price of more than one good.]

3.I.11[B] Suppose that we know not only p^0, p^1, and x^0 but also $x^1 = x(p^1, w)$. Show that if $(p^1 - p^0)\cdot x^1 > 0$, then the consumer must be worse off at price–wealth situation (p^1, w) than at (p^0, w). Interpret this test as a first-order approximation to the expenditure function at p^1.

Also show that an alternative way to write this test is $p^0 \cdot (x^1 - x^0) < 0$, and depict the test for the case where $L = 2$ in (x_1, x_2) space. [*Hint*: Locate the point x^0 on the set $\{x \in \mathbb{R}^L_+ : u(x) = u^0\}$.]

3.I.12$^{\mathrm{B}}$ Extend the compensating and equivalent variation measures of welfare change to the case of changes in both prices and wealth, so that we change from (p^0, w^0) to (p^1, w^1). Also extend the "partial information" test developed in Section 3.I to this case.

3.J.1$^{\mathrm{C}}$ Show that when $L = 2$, $x(p, w)$ satisfies the strong axiom if and only if it satisfies the weak axiom.

3.AA.1$^{\mathrm{B}}$ Suppose that the consumption set is $X = \{x \in \mathbb{R}^2_+ : x_1 + x_2 \geq 1\}$ and the utility function is $u(x) = x_2$. Represent graphically, and show (a) that the locally cheaper consumption test fails at $(p, w) = (1, 1, 1)$ and (b) that market demand is not continuous at this point. Interpet economically.

3.AA.2$^{\mathrm{C}}$ Under the conditions of Proposition 3.AA.1, show that $h(p, u)$ is upper hemicontinuous and that $e(p, u)$ is continuous (even if we replace minimum by infimum and allow $p \geq 0$). Also, assuming that $h(p, u)$ is a function, give conditions for its differentiability.

CHAPTER

4

Aggregate Demand

4.A Introduction

For most questions in economics, the aggregate behavior of consumers is more important than the behavior of any single consumer. In this chapter, we investigate the extent to which the theory presented in Chapters 1 to 3 can be applied to *aggregate demand*, a suitably defined sum of the demands arising from all the economy's consumers. There are, in fact, a number of different properties of individual demand that we might hope would also hold in the aggregate. Which ones we are interested in at any given moment depend on the particular application at hand.

In this chapter, we ask three questions about aggregate demand:

(i) Individual demand can be expressed as a function of prices and the individual's wealth level. *When can aggregate demand be expressed as a function of prices and aggregate wealth?*

(ii) Individual demand derived from rational preferences necessarily satisfies the weak axiom of revealed preference. *When does aggregate demand satisfy the weak axiom?* More generally, when can we apply in the aggregate the demand theory developed in Chapter 2 (especially Section 2.F)?

(iii) Individual demand has welfare significance; from it, we can derive measures of welfare change for the consumer, as discussed in Section 3.I. *When does aggregate demand have welfare significance?* In particular, when do the welfare measures discussed in Section 3.I have meaning when they are computed from the aggregate demand function?

These three questions could, with a grain of salt, be called the *aggregation theories of*, respectively, *the econometrician*, *the positive theorist*, and *the welfare theorist*.

The econometrician is interested in the degree to which he can impose a simple structure on aggregate demand functions in estimation procedures. One aspect of these concerns, which we address here, is the extent to which aggregate demand can be accurately modeled as a function of only *aggregate* variables, such as aggregate (or, equivalently, average) consumer wealth. This question is important because the econometrician's data may be available only in an aggregate form.

The positive (behavioral) theorist, on the other hand, is interested in the degree

to which the positive restrictions of individual demand theory apply in the aggregate. This can be significant for deriving predictions from models of market equilibrium in which aggregate demand plays a central role.[1]

The welfare theorist is interested in the normative implications of aggregate demand. He wants to use the measures of welfare change derived in Section 3.I to evaluate the welfare significance of changes in the economic environment. Ideally, he would like to treat aggregate demand as if it were generated by a "representative consumer" and use the changes in this fictional individual's welfare as a measure of aggregate welfare.

Although the conditions we identify as important for each of these aggregation questions are closely related, the questions being asked in the three cases are conceptually quite distinct. Overall, we shall see that, in all three cases, very strong restrictions will need to hold for the desired aggregation properties to obtain. We discuss these three questions, in turn, in Sections 4.B to 4.D.

Finally, Appendix A discusses the regularizing (i.e., "smoothing") effects arising from aggregation over a large number of consumers.

4.B Aggregate Demand and Aggregate Wealth

Suppose that there are I consumers with rational preference relations $\succsim_i$ and corresponding Walrasian demand functions $x_i(p, w_i)$. In general, given prices $p \in \mathbb{R}^L$ and wealth levels $(w_1, \ldots, w_I)$ for the I consumers, aggregate demand can be written as

$$x(p, w_1, \ldots, w_I) = \sum_{i=1}^{I} x_i(p, w_i).$$

Thus, aggregate demand depends not only on prices but also on the specific wealth levels of the various consumers. In this section, we ask when we are justified in writing aggregate demand in the simpler form $x(p, \sum_i w_i)$, where aggregate demand depends only on aggregate wealth $\sum_i w_i$.

For this property to hold in all generality, aggregate demand must be identical for any two distributions of the same total amount of wealth across consumers. That is, for any $(w_1, \ldots, w_I)$ and $(w'_1, \ldots, w'_I)$ such that $\sum_i w_i = \sum_i w'_i$, we must have $\sum_i x_i(p, w_i) = \sum_i x_i(p, w'_i)$.

To examine when this condition is satisfied, consider, starting from some initial distribution $(w_1, \ldots, w_I)$, a differential change in wealth $(dw_1, \ldots, dw_I) \in \mathbb{R}^I$ satisfying $\sum_i dw_i = 0$. If aggregate demand can be written as a function of aggregate wealth, then assuming differentiability of the demand functions, we must have

$$\sum_i \frac{\partial x_{\ell i}(p, w_i)}{\partial w_i} dw_i = 0 \quad \text{for every } \ell.$$

This can be true for all redistributions $(dw_1, \ldots, dw_I)$ satisfying $\sum_i dw_i = 0$ and from any initial wealth distribution $(w_1, \ldots, w_I)$ if and only if the coefficients of the different

1. The econometrician may also be interested in these questions because a priori restrictions on the properties of aggregate demand can be incorporated into his estimation procedures.

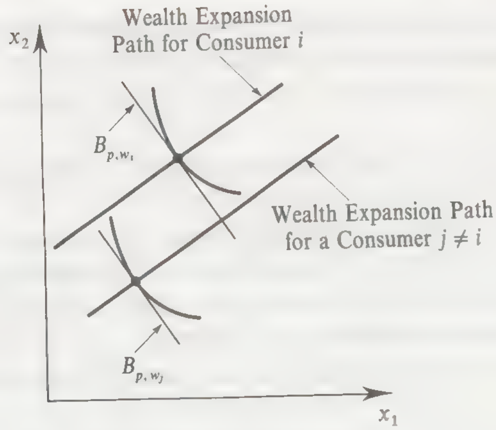


Figure 4.B.1
Invariance of aggregate demand to redistribution of wealth implies wealth expansion paths that are straight and parallel across consumers.

dw_i are equal; that is,

$$\frac{\partial x_{\ell i}(p, w_i)}{\partial w_i} = \frac{\partial x_{\ell j}(p, w_j)}{\partial w_j} \tag{4.B.1}$$

for every ℓ, any two individuals i and j, and all $(w_1, \ldots, w_I)$.[2]

In short, for any fixed price vector p, and any commodity ℓ, the wealth effect at p must be the same whatever consumer we look at and whatever his level of wealth.[3] It is indeed fairly intuitive that in this case, the individual demand changes arising from any wealth redistribution across consumers will cancel out. Geometrically, the condition is equivalent to the statement that all consumers' wealth expansion paths are parallel, straight lines. Figure 4.B.1 depicts parallel, straight wealth expansion paths.

One special case in which this property holds arises when all consumers have identical preferences that are homothetic. Another is when all consumers have preferences that are quasilinear with respect to the same good. Both cases are examples of a more general result shown in Proposition 4.B.1.

Proposition 4.B.1: A necessary and sufficient condition for the set of consumers to exhibit parallel, straight wealth expansion paths at any price vector p is that preferences admit indirect utility functions of the Gorman form with the coefficients on w_i the same for every consumer i. That is:

$$v_i(p, w_i) = a_i(p) + b(p) w_i.$$

Proof: You are asked to establish sufficiency in Exercise 4.B.1 (this is not too difficult; use Roy's identity). Keep in mind that we are neglecting boundaries (alternatively, the significance of a result such as this is only local). You should not attempt to prove necessity. For a discussion of this result, see Deaton and Muellbauer (1980). ■

---

2. As usual, we are neglecting boundary constraints; hence, strictly speaking, the validity of our claims in this section is only local.

3. Note that $\partial x_{\ell i}(p, w_i)/\partial w_i = \partial x_{\ell i}(p, w_i')/\partial w_i$ for all $w_i \neq w_i'$ because for any values of $w_j, j \neq i$, (4.B.1) must hold for the wealth distributions $(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_I)$ and $(w_1, \ldots, w_{i-1}, w_i', w_{i+1}, \ldots, w_I)$. Hence, $\partial x_{\ell i}(p, w_i)/\partial w_i = \partial x_{\ell j}(p, w_j)/\partial w_j = \partial x_{\ell i}(p, w_i')/\partial w_i$ for any $j \neq i$.

Thus, aggregate demand can be written as a function of aggregate wealth if and only if all consumers have preferences that admit indirect utility functions of the Gorman form with equal wealth coefficients $b(p)$. Needless to say, this is a very restrictive condition on preferences.[4]

Given this conclusion, we might ask whether less restrictive conditions can be obtained if we consider aggregate demand functions that depend on a wider set of aggregate variables than just the total (or, equivalently, the mean) wealth level. For example, aggregate demand might be allowed to depend on both the mean and the variance of the statistical distribution of wealth or even on the whole statistical distribution itself. Note that the latter condition is still restrictive. It implies that aggregate demand depends only on how many rich and poor there are, not on who in particular is rich or poor.

These more general forms of dependence on the distribution of wealth are indeed valid under weaker conditions than those required for aggregate demand to depend only on aggregate wealth. For a trivial example, note that aggregate demand depends only on the statistical distribution of wealth whenever all consumers possess identical but otherwise arbitrary preferences and differ only in their wealth levels. We shall not pursue this topic further here; good references are Deaton and Muellbauer (1980), Lau (1982) and Jorgenson (1990).

There is another way in which we might be able to get a more positive answer to our question. So far, the test that we have applied is whether the aggregate demand function can be written as a function of aggregate wealth for *any* distribution of wealth across consumers. The requirement that this be true for every conceivable wealth distribution is a strong one. Indeed, in many situations, individual wealth levels may be generated by some underlying process that restricts the set of individual wealth levels which can arise. If so, it may still be possible to write aggregate demand as a function of prices and aggregate wealth.

For example, when we consider general equilibrium models in Part IV, individual wealth is generated by individuals' shareholdings of firms and by their ownership of given, fixed stocks of commodities. Thus, the individual levels of real wealth are determined as a function of the prevailing price vector.

Alternatively, individual wealth levels may be determined in part by various government programs that redistribute wealth across consumers (see Section 4.D). Again, these programs may limit the set of possible wealth distributions that may arise.

To see how this can help, consider an extreme case. Suppose that individual i's wealth level is generated by some process that can be described as a function of prices p and aggregate wealth w, $w_i(p, w)$. This was true, for example, in the general equilibrium illustration above. Similarly, the government program may base an individual's taxes (and hence his final wealth position) on his wage rate and the total (real) wealth of the society. We call a family of functions $(w_1(p, w), \dots, w_I(p, w))$ with $\sum_i w_i(p, w) = w$ for all (p, w) a *wealth distribution rule*. When individual wealth levels

4. Recall, however, that it includes some interesting and important classes of preferences. For example, if preferences are quasilinear with respect to good ℓ, then there is an indirect utility of the form $a_i(p) + w_i/p_\ell$, which, letting $b(p) = 1/p_\ell$, we can see is of the Gorman type with identical $b(p)$.

are generated by a wealth distribution rule, we can indeed *always* write aggregate demand as a function $x(p, w) = \sum_i x_i(p, w_i(p, w))$, and so aggregate demand depends only on prices and aggregate wealth.

4.C Aggregate Demand and the Weak Axiom

To what extent do the positive properties of individual demand carry over to the aggregate demand function $x(p, w_1, \dots, w_I) = \sum_i x_i(p, w_i)$? We can note immediately three properties that do: continuity, homogeneity of degree zero, and Walras' law [that is, $p \cdot x(p, w_1, \dots, w_I) = \sum_I w_i$ for all $(p, w_1, \dots, w_I)$]. In this section, we focus on the conditions under which aggregate demand also satisfies the weak axiom, arguably the most central positive property of the individual Walrasian demand function.

To study this question, we would like to operate on an aggregate demand written in the form $x(p, w)$, where w is aggregate wealth. This is the form for which we gave the definition of the weak axiom in Chapter 2. We accomplish this by supposing that there is a wealth distribution rule $(w_1(p, w), \dots, w_I(p, w))$ determining individual wealths from the price vector and total wealth. We refer to the end of Section 4.B for a discussion of wealth distribution rules.[5] With the wealth distribution rule at our disposal, aggregate demand can automatically be written as

$$x(p, w) = \sum_i x_i(p, w_i(p, w)).$$

Formally, therefore, the aggregate demand function $x(p, w)$ depends then only on aggregate wealth and is therefore a market demand function in the sense discussed in Chapter 2.[6] We now investigate the fulfillment of the weak axiom by $x(\cdot, \cdot)$.

In point of fact, and merely for the sake of concreteness, we shall be even more specific and focus on a particularly simple example of a distribution rule. Namely, we restrict ourselves to the case in which relative wealths of the consumers remain fixed, that is, are independent of prices. Thus, we assume that we are given wealth shares $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$, so that $w_i(p, w) = \alpha_i w$ for every level $w \in \mathbb{R}$ of aggregate wealth.[7] We have then

$$x(p, w) = \sum_i x_i(p, \alpha_i w).$$

We begin by recalling from Chapter 2 the definition of the weak axiom.

Definition 4.C.1: The aggregate demand function $x(p, w)$ satisfies the weak axiom (WA) if $p \cdot x(p', w') \leq w$ and $x(p, w) \neq x(p', w')$ imply $p' \cdot x(p, w) > w'$ for any (p, w) and (p', w').

---

5. There is also a methodological advantage to assuming the presence of a wealth distribution rule. It avoids confounding different aggregation issues because the aggregation problem studied in Section 4.B (invariance of demand to redistributions) is then entirely assumed away.

6. Note that it assigns commodity bundles to price–wealth combinations, and, provided every $w_i(\cdot, \cdot)$ is continuous and homogeneous of degree one, that it is continuous, homogeneous of degree zero, and satisfies Walras's law.

7. Observe that this distribution rule amounts to leaving the wealth levels $(w_1, \dots, w_I)$ unaltered and considering only changes in the price vector p. This is because the homogeneity of degree zero of $x(p, w_1, \dots, w_I)$ implies that any proportional change in wealths can also be captured by a proportional change in prices. The description by means of shares is, however, analytically more convenient.

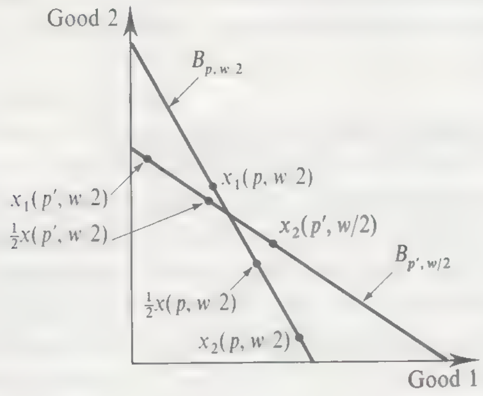


Figure 4.C.1
Failure of aggregate demand to satisfy the weak axiom.

We next provide an example illustrating that aggregate demand may not satisfy the weak axiom.

Example 4.C.1: *Failure of Aggregate Demand to Satisfy the WA.* Suppose that there are two commodities and two consumers. Wealth is distributed equally so that $w_1 = w_2 = w/2$, where w is aggregate wealth. Two price vectors p and p' with corresponding individual demands $x_1(p, w/2)$ and $x_2(p, w/2)$ under p, and $x_1(p', w/2)$ and $x_2(p', w/2)$ under p', are depicted in Figure 4.C.1.

These individual demands satisfy the weak axiom, but the aggregate demands do not. Figure 4.C.1 shows the vectors $\frac{1}{2}x(p, w)$ and $\frac{1}{2}x(p', w)$, which are equal to the average of the two consumers' demands; (and so for each price vector, they must lie at the midpoint of the line segment connecting the two individuals' consumption vectors). As illustrated in the figure, we have

$$\tfrac{1}{2}p\cdot x(p', w) < w/2 \qquad \text{and} \qquad \tfrac{1}{2}p'\cdot x(p, w) < w/2,$$

which (multiply both sides by 2) constitutes a violation of the weak axiom at the price-wealth pairs considered. ■

The reason for the failure illustrated in Example 4.C.1 can be traced to wealth effects. Recall from Chapter 2 (Proposition 2.F.1) that $x(p, w)$ satisfies the weak axiom if and only if it satisfies the law of demand for *compensated* price changes. Precisely, if and only if for any (p, w) and any price change p' that is compensated [so that $w' = p'\cdot x(p, w)$], we have

$$(p' - p)\cdot[x(p', w') - x(p, w)] \leq 0, \tag{4.C.1}$$

with strict inequality if $x(p, w) \neq x(p', w')$.[8]

If the price–wealth change under consideration, say from (p, w) to (p', w'), happened to be a compensated price change for *every* consumer i—that is, if $\alpha_i w' = p'\cdot x_i(p, \alpha_i w)$ for all i—then because individual demand satisfies the weak axiom, we would know (again by Proposition 2.F.1) that for all $i = 1, \ldots, I$:

$$(p' - p)\cdot[x_i(p', \alpha_i w') - x_i(p, \alpha_i w)] \leq 0, \tag{4.C.2}$$

8. Note that if $p\cdot x(p', w') \leq w$ and $x(p', w') \neq x(p, w)$, then we must have $p'\cdot x(p, w) > w'$, in agreement with the weak axiom.

with strict inequality if $x_i(p', \alpha_i w) \neq x_i(p, \alpha_i w')$. Adding (4.C.2) over i gives us precisely (4.C.1). Thus, we conclude that aggregate demand must satisfy the WA for any price–wealth change that is compensated for every consumer.

The difficulty arises because a price-wealth change that is compensated in the aggregate, so that $w' = p' \cdot x(p, w)$, need not be compensated for each individual; we may well have $\alpha_i w' \neq p' \cdot x_i(p, \alpha_i w)$ for some or all i. If so, the individual wealth effects [which, except for the condition $p \cdot D_{w_i} x(p, \alpha_i w) = 1$, are essentially unrestricted] can play havoc with the well-behaved but possibly small individual substitution effects. The result may be that (4.C.2) fails to hold for some i, thus making possible the failure of the similar expression (4.C.1) in the aggregate.

Given that a property of individual demand as basic as the WA cannot be expected to hold generally for aggregate demand, we might wish to know whether there are any restrictions on individual preferences under which it must be satisfied. The preceding discussion suggests that it may be worth exploring the implications of assuming that the law of demand, expression (4.C.2), holds at the individual level for price changes that are left uncompensated. Suppose, indeed, that given an initial position (p, w_i), we consider a price change p' that is not compensated, namely, we leave $w_i' = w_i$. If (4.C.2) nonetheless holds, then by addition so does (4.C.1). More formally, we begin with a definition.

Definition 4.C.2: The individual demand function $x_i(p, w_i)$ satisfies the *uncompensated law of demand* (*ULD*) property if

$$(p' - p) \cdot [x_i(p', w_i) - x_i(p, w_i)] \leq 0 \qquad (4.C.3)$$

for any p, p', and w_i, with strict inequality if $x_i(p', w_i) \neq x_i(p, w_i)$. The analogous definition applies to the aggregate demand function $x(p, w)$.

In view of our discussion of the weak axiom in Section 2.F, the following differential version of the ULD property should come as no surprise (you are asked to prove it in Exercise 4.C.1):

> If $x_i(p, w_i)$ satisfies the ULD property, then $D_p x_i(p, w_i)$ is negative semidefinite; that is, $dp \cdot D_p x_i(p, w_i)\, dp \leq 0$ for all dp.

As with the weak axiom, there is a converse to this:

> If $D_p x_i(p, w_i)$ is negative definite for all p, then $x_i(p, w_i)$ satisfies the ULD property.

The analogous differential version holds for the aggregate demand function $x(p, w)$.

The great virtue of the ULD property is that, in contrast with the WA, it does, in fact, aggregate. Adding the individual condition (4.C.3) for $w_i = \alpha_i w$ gives us $(p' - p) \cdot [x(p', w) - x(p, w)] \leq 0$, with strict inequality if $x(p, w) \neq x'(p, w)$. This leads us to Proposition 4.C.1.

Proposition 4.C.1: If every consumer's Walrasian demand function $x_i(p, w_i)$ satisfies the uncompensated law of demand (ULD) property, so does the aggregate demand $x(p, w) = \sum_i x_i(p, \alpha_i w)$. As a consequence, the aggregate demand $x(p, w)$ satisfies the weak axiom.

Proof: Consider any (p, w), (p', w) with $x(p, w) \neq x(p', w)$. We must have

$$x_i(p, \alpha_i w) \neq x_i(p', \alpha_i w)$$

for some i. Therefore, adding (4.C.3) over i, we get

$$(p' - p) \cdot [x(p, w) - x(p', w)] < 0.$$

This holds for all p, p', and w.

To verify the WA, take any (p, w), (p', w') with $x(p, w) \neq x(p', w')$ and $p \cdot x(p', w') \leq w$.[9] Define $p'' = (w/w')p'$. By homogeneity of degree zero, we have $x(p'', w) = x(p', w')$. From $(p'' - p) \cdot [x(p'', w) - x(p, w)] < 0$, $p \cdot x(p'', w) \leq w$, and Walras' law, it follows that $p'' \cdot x(p, w) > w$. That is, $p' \cdot x(p, w) > w'$. ■

How restrictive is the ULD property as an axiom of individual behavior? It is clearly not implied by preference maximization (see Exercise 4.C.3). Propositions 4.C.2 and 4.C.3 provide sufficient conditions for individual demands to satisfy the ULD property.

Proposition 4.C.2: If $\succsim_i$ is homothetic, then $x_i(p, w_i)$ satisfies the uncompensated law of demand (ULD) property.

Proof: We consider the differentiable case [i.e., we assume that $x_i(p, w_i)$ is differentiable and that $\succsim_i$ is representable by a differentiable utility function]. The matrix $D_p x_i(p, w_i)$ is

$$D_p x_i(p, w_i) = S_i(p, w_i) - \frac{1}{w_i} x_i(p, w_i)\, x_i(p, w_i)^T,$$

where $S_i(p, w_i)$ is consumer i's Slutsky matrix. Because $[dp \cdot x_i(p, w_i)]^2 > 0$ except when $dp \cdot x_i(p, w_i) = 0$ and $dp \cdot S_i(p, w_i)\, dp < 0$ except when dp is proportional to p, we can conclude that $D_p x_i(p, w_i)$ is negative definite, and so the ULD condition holds. ■

In Proposition 4.C.2, the conclusion is obtained with minimal help from the substitution effects. Those could all be arbitrarily small. The wealth effects by themselves turn out to be sufficiently well behaved. Unfortunately, the homothetic case is the only one in which this is so (see Exercise 4.C.4). More generally, for the ULD property to hold, the substitution effects (which are always well behaved) must be large enough to overcome possible "perversities" coming from the wealth effects. The intriguing result in Proposition 4.3.C [due to Mitiushin and Polterovich (1978) and Milleron (1974); see Mas-Colell (1991) for an account and discussion of this result] gives a concrete expression to this relative dominance of the substitution effects.

Proposition 4.C.3: Suppose that $\succsim_i$ is defined on the consumption set $X = \mathbb{R}^L_+$ and is representable by a twice continuously differentiable concave function $u_i(\cdot)$. If

$$-\frac{x_i \cdot D^2 u_i(x_i) x_i}{x_i \cdot \nabla u_i(x_i)} < 4 \quad \text{for all } x_i,$$

then $x_i(p, w_i)$ satisfies the unrestricted law of demand (ULD) property.

9. Strictly speaking, this proof is required because although we know that the WA is equivalent to the law of demand for compensated price changes, we are now dealing with uncompensated price changes.

The proof of Proposition 4.C.3 will not be given. The courageous reader can attempt it in Exercise 4.C.5.

The condition in Proposition 4.C.3 is not an extremely stringent one. In particular, notice how amply the homothetic case fits into it (Exercise 4.C.6). So, to the question "How restrictive is the ULD property as an axiom of individual behavior?" perhaps we can answer: "restrictive, but not extremely so."[10]

Note, in addition, that for the ULD property to hold for aggregate demand, it is not necessary that the ULD be satisfied at the individual level. It may arise out of aggregation itself. The example in Proposition 4.C.4, due to Hildenbrand (1983), is not very realistic, but it is nonetheless highly suggestive.

Proposition 4.C.4: Suppose that all consumers have identical preferences $\succsim$ defined on $\mathbb{R}^L_+$ [with individual demand functions denoted $\tilde{x}(p, w)$] and that individual wealth is uniformly distributed on an interval $[0, \bar{w}]$ (strictly speaking, this requires a continuum of consumers). Then the aggregate (rigorously, the average) demand function

$$x(p) = \int_0^{\bar{w}} \tilde{x}(p, w)\, dw$$

satisfies the unrestricted law of demand (ULD) property.

Proof: Consider the differentiable case. Take $v \neq 0$. Then

$$v \cdot Dx(p)v = \int_0^{\bar{w}} v \cdot D_p \tilde{x}(p, w)v\, dw.$$

Also

$$D_p \tilde{x}(p, w) = S(p, w) - D_w \tilde{x}(p, w)\tilde{x}(p, w)^T,$$

where $S(p, w)$ is the Slutsky matrix of the individual demand function $x(\cdot, \cdot)$ at (p, w). Hence,

$$v \cdot Dx(p)v = \int_0^{\bar{w}} v \cdot S(p, w)v\, dw - \int_0^{\bar{w}} (v \cdot D_w \tilde{x}(p, w))(v \cdot \tilde{x}(p, w))\, dw.$$

The first term of this sum is negative, unless v is proportional to p. For the second, note that

$$2(v \cdot D_w \tilde{x}(p, w))(v \cdot \tilde{x}(p, w)) = \frac{d(v \cdot \tilde{x}(p, w))^2}{dw}.$$

So

$$-\int_0^{\bar{w}} (v \cdot D_w \tilde{x}(p, w))(v \cdot \tilde{x}(p, w))\, dw = -\frac{1}{2}\int_0^{\bar{w}} \frac{d(v \cdot \tilde{x}(p, w))^2}{dw}\, dw = -\frac{1}{2}(v \cdot \tilde{x}(p, \bar{w}))^2 \leq 0,$$

where we have used $\tilde{x}(p, 0) = 0$. Observe that the sign is negative when v is proportional to p. ■

Recall that the ULD property is additive across groups of consumers. Therefore, what we need in order to apply Proposition 4.C.4 is, not that preferences be identical, but that for every preference relation, the distribution of wealth conditional on that preference be uniform over

---

10. Not to misrepresent the import of this claim, we should emphasize that Proposition 4.C.1, which asserts that the ULD property is preserved under addition, holds for the price-independent distribution rules that we are considering in this section. When the distribution of real wealth may depend on prices (as it typically will in the general equilibrium applications of Part IV), then aggregate demand may violate the WA even if individual demand satisfies the ULD property (see Exercise 4.C.13). We discuss this point further in Section 17.F.

some interval that includes the level 0 (in fact, a nonincreasing density function is enough; see Exercise 4.C.7).

One lesson of Proposition 4.C.4 is that the properties of aggregate demand will depend on how preferences and wealth are distributed. We could therefore pose the problem quite generally and ask which distributional conditions on preferences and wealth will lead to satisfaction of the weak axiom by aggregate demand.[11]

As mentioned in Section 2.F, a market demand function $x(p, w)$ can be shown to satisfy the WA if for all (p, w), the Slutsky matrix $S(p, w)$ derived from the function $x(p, w)$ satisfies $dp \cdot S(p, w)\, dp < 0$ for every $dp \neq 0$ not proportional to p. We now examine when this property might hold for the aggregate demand function.

The Slutsky equation for the aggregate demand function is

$$S(p, w) = D_p x(p, w) + D_w x(p, w)\, x(p, w)^{\mathrm{T}}. \tag{4.C.4}$$

Or, since $x(p, w) = \sum_i x_i(p, \alpha_i w)$,

$$S(p, w) = D_p x(p, w) + \left[\sum_i \alpha_i D_{w_i} x_i(p, \alpha_i w)\right] x(p, w)^{\mathrm{T}} \tag{4.C.5}$$

Next, let $S_i(p, w_i)$ denote the individual Slutsky matrices. Adding the individual Slutsky equations gives

$$\sum_i S_i(p, \alpha_i w) = \sum_i D_p x_i(p, \alpha_i w) + \sum_i D_{w_i} x_i(p, \alpha_i w)\, x_i(p, \alpha_i w)^{\mathrm{T}} \tag{4.C.6}$$

Since $D_p x(p, w) = \sum_i D_p x_i(p, \alpha_i w)$, we can substitute (4.C.6) into (4.C.5) to get

$$S(p, w) = \sum_i S_i(p, w_i) - \sum_i \alpha_i [D_{w_i} x_i(p, \alpha_i w) - D_w x(p, w)] \left[\frac{1}{\alpha_i} x_i(p, \alpha_i w) - x(p, w)\right]^{\mathrm{T}}. \tag{4.C.7}$$

Note that because of wealth effects, the Slutsky matrix of aggregate demand is *not* the sum of the individual Slutsky matrices. The difference

$$\begin{aligned} C(p, w) &= \sum_i S_i(p, \alpha w) - S(p, w) \\ &= \sum_i \alpha_i [D_{w_i} x_i(p, \alpha_i w) - D_w x(p, w)] \left[\frac{1}{\alpha_i} x_i(p, \alpha_i w) - x(p, w)\right]^{\mathrm{T}} \end{aligned} \tag{4.C.8}$$

is a covariance matrix between wealth effect vectors $D_{w_i} x_i(p, \alpha_i w)$ and proportionately adjusted consumption vectors $(1/\alpha_i) x_i(p, \alpha_i w)$. The former measures how the marginal dollar is spent across commodities; the latter measures the same thing for the average dollar [e.g., $(1/\alpha_i w) x_{\ell i}(p, \alpha_i w)$ is the per-unit-of-wealth consumption of good ℓ by consumer i]. Every "observation" receives weight α_i. Note also that, as it should be, we have

$$\sum_i \alpha_i [D_{w_i} x_i(p, \alpha_i w) - D_w x(p, w)] = 0 \quad \text{and} \quad \sum_i \alpha_i [(1/\alpha_i) x_i(p, \alpha_i w) - x(p, w)] = 0.$$

For an individual Slutsky matrix $S_i(\cdot, \cdot)$ we always have $dp \cdot S_i(p, \alpha_i w)\, dp < 0$ for $dp \neq 0$ not proportional to p. Hence, a *sufficient* condition for the Slutsky matrix of aggregate demand to have the desired property is that $C(p, w)$ be positive semidefinite. Speaking loosely, this will be the case if, on average, there is a *positive association* across consumers between consumption (per unit of wealth) in one commodity and the wealth effect for that commodity.

Figure 4.C.2(a) depicts a case for $L = 2$ in which, assuming a uniform distribution of wealth across consumers, this association is positive: Consumers with higher-than-average

11. In the next few paragraphs, we follow Jerison (1982) and Freixas and Mas-Colell (1987).

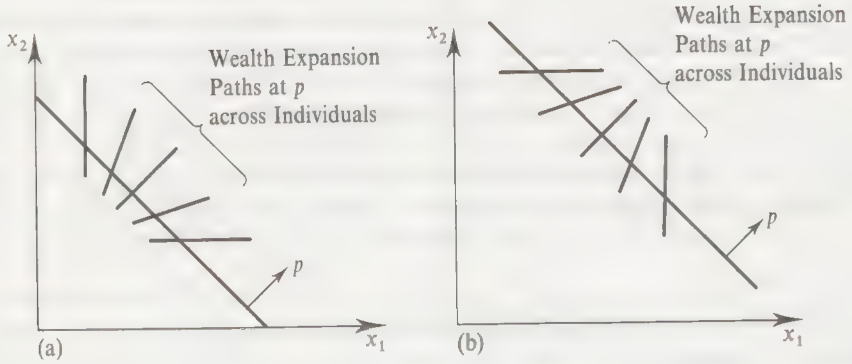


Figure 4.C.2
The relation across consumers between expenditure per unit of wealth on a commodity and its wealth effect when all consumers have the same wealth.
(a) Positive relation.
(b) Negative relation.

consumption of one good spend a higher-than-average fraction of their last unit of wealth on that good. The association is negative in Figure 4.C.2(b).[12,13]

From the preceding derivation, we can see that aggregate demand satisfies the WA in two cases of interest: (i) All the $D_{w_i}x_i(p, \alpha_i w)$ are equal (there are equal wealth effects), and (ii) all the $(1/\alpha_i)x_i(p, \alpha_i w)$ are equal (there is proportional consumption). In both cases, we have $C(p, w) = 0$, and so $dp \cdot S(p, w)\, dp < 0$ whenever $dp \neq 0$ is not proportional to p.

Case (i) has important implications. In particular, if every consumer has indirect utility functions of the Gorman form $v_i(p, w_i) = a_i(p) + b(p)w$, with the coefficient $b(p)$ identical across consumers, then (as we saw in Section 4.B) the wealth effects are the same for all consumers and we can therefore conclude that the WA is satisfied. We know from Section 4.B that one is led to this family of indirect utility functions by the requirement that aggregate demand be invariant to redistribution of wealth. Thus, aggregate demand satisfying the weak axiom for a fixed distribution of wealth is a less demanding property than the invariance to redistribution property considered in Section 4.B. In particular, if the second property holds, then the first also holds, but aggregate demand (for a fixed distribution of wealth) may satisfy the weak axiom even though aggregate demand may not be invariant to redistribution of wealth (e.g., individual preferences may be homothetic but not identical).

Having spent all this time investigating the weak axiom (WA), you might ask: "What about the strong axiom (SA)?" We have not focused on the Strong Axiom for three reasons.

First, the WA is a robust property, whereas the SA (which, remember, yields the symmetry of the Slutsky matrix) is not; a priori, the chances of it being satisfied by a real economy are essentially zero. For example, if we start with a group of consumers with identical preferences and wealth, then aggregate demand obviously satisfies the SA. However, if we now perturb every preference slightly and independently across consumers, the negative semidefiniteness of the Slutsky matrices (and therefore the WA) may well be preserved but the symmetry (and therefore the SA) will almost certainly not be.

12. You may want to verify that the wealth expansion paths of Example 4.C.1 must indeed look like Figure 4.C.2(b).

13. A priori, we cannot say which form is more likely. Because the demand at zero wealth is zero, it is true that for a consumer, *some* dollar must be spent among the two goods according to shares similar to the shares of the average dollar. But if the levels of wealth are not close to zero, it does not follow that this is the case for the *marginal* dollar. It may even happen that because of incipient satiation, the shares of the marginal dollar display consumption propensities that are the reverse of the ones exhibited by the average dollar. See Hildenbrand (1994) for an account of empirical research on this matter.

Second, many of the strong positive results of general equilibrium (to be reviewed in Part IV, especially Chapters 15 and 17) to which one wishes to apply the aggregation theory discussed in this chapter depend on the weak axiom, not on the strong axiom, holding in the aggregate.

Third, while one might initially think that the existence of a preference relation explaining aggregate behavior (which is what we get from the SA) would be the condition required to use aggregate demand measures (such as aggregate consumer surplus) as welfare indicators, we will see in Section 4.D that, in fact, more than this condition is required anyway.

4.D Aggregate Demand and the Existence of a Representative Consumer

The aggregation question we pose in this section is: When can we compute meaningful measures of aggregate welfare using the aggregate demand function and the welfare measurement techniques discussed in Section 3.I for individual consumers? More specifically, when can we treat the aggregate demand function as if it were generated by a fictional *representative consumer* whose preferences can be used as a measure of aggregate societal (or *social*) welfare?

We take as our starting point a distribution rule $(w_1(p, w), \dots, w_I(p, w))$ that to every level of aggregate wealth $w \in \mathbb{R}$ assigns individual wealths. We assume that $\sum_i w_i(p, w) = w$ for all (p, w) and that every $w_i(\cdot, \cdot)$ is continuous and homogeneous of degree one. As discussed in Sections 4.B and 4.C, aggregate demand then takes the form of a conventional market demand function $x(p, w) = \sum_i x_i(p, w_i(p, w))$. In particular, $x(p, w)$ is continuous, is homogeneous of degree zero, and satisfies Walras' law. It is important to keep in mind that the aggregate demand function $x(p, w)$ depends on the wealth distribution rule (except under the special conditions identified in Section 4.B).

It is useful to begin by distinguishing two senses in which we could say that there is a representative consumer. The first is a positive, or behavioral, sense.

Definition 4.D.1: A *positive representative consumer* exists if there is a rational preference relation $\succsim$ on $\mathbb{R}^L_+$ such that the aggregate demand function $x(p, w)$ is precisely the Walrasian demand function generated by this preference relation. That is, $x(p, w) \succ x$ whenever $x \neq x(p, w)$ and $p \cdot x \leq w$.

A positive representative consumer can thus be thought of as a fictional individual whose utility maximization problem when facing society's budget set $\{x \in \mathbb{R}^L_+ : p \cdot x \leq w\}$ would generate the economy's aggregate demand function.

For it to be correct to treat aggregate demand as we did individual demand functions in Section 3.I, there must be a positive representative consumer.[14] However, although this is a necessary condition for the property of aggregate demand that we seek, it is not sufficient. We also need to be able to assign welfare significance to this

14. Note that if there is a positive representative consumer, then aggregate demand satisfies the positive properties sought in Section 4.C. Indeed, not only will aggregate demand satisfy the weak axiom, but it will also satisfy the strong axiom. Thus, the aggregation property we are after in this section is stronger than the one discussed in Section 4.C.

fictional individual's demand function. This will lead to the definition of a *normative* representative consumer. To do so, however, we first have to be more specific about what we mean by the term *social welfare*. We accomplish this by introducing the concept of a *social welfare function*, a function that provides a summary (social) utility index for any collection of individual utilities.

Definition 4.D.2: A (*Bergson-Samuelson*) *social welfare function* is a function $W: \mathbb{R}^I \to \mathbb{R}$ that assigns a utility value to each possible vector $(u_1, \dots, u_I) \in \mathbb{R}^I$ of utility levels for the I consumers in the economy.

The idea behind a social welfare function $W(u_1, \dots, u_I)$ is that it accurately expresses society's judgments on how individual utilities have to be compared to produce an ordering of possible social outcomes. (We do not discuss in this section the issue of where this social preference ranking comes from. Chapters 21 and 22 cover this point in much more detail.) We also assume that social welfare functions are increasing, concave, and whenever convenient, differentiable.

Let us now hypothesize that there is a process, a benevolent central authority perhaps, that, for any given prices p and aggregate wealth level w, redistributes wealth in order to maximize social welfare. That is, for any (p, w), the wealth distribution $(w_1(p, w), \dots, w_I(p, w))$ solves

$$\begin{aligned} \underset{w_1, \dots, w_I}{\text{Max}} \quad & W(v_1(p, w_1), \dots, v_I(p, w_I)) \\ \text{s.t.} \ & \textstyle\sum_{i=1}^{I} w_i \leq w, \end{aligned} \tag{4.D.1}$$

where $v_i(p, w)$ is consumer i's indirect utility function.[15,16] The optimum value of problem (4.D.1) defines a social indirect utility function $v(p, w)$. Proposition 4.D.1 shows that this indirect utility function provides a positive representative consumer for the aggregate demand function $x(p, w) = \sum_i x_i(p, w_i(p, w))$.

Proposition 4.D.1: Suppose that for each level of prices p and aggregate wealth w, the wealth distribution $(w_1(p, w), \dots, w_I(p, w))$ solves problem (4.D.1). Then the value function $v(p, w)$ of problem (4.D.1) is an indirect utility function of a positive representative consumer for the aggregate demand function $x(p, w) = \sum_i x_i(p, w_i(p, w))$.

Proof: In Exercise 4.D.2, you are asked to establish that $v(p, w)$ does indeed have the properties of an indirect utility function. The argument for the proof then consists of using Roy's identity to derive a Walrasian demand function from $v(p, w)$, which we denote by $x_R(p, w)$, and then establishing that it actually equals $x(p, w)$.

We begin by recording the first-order conditions of problem (4.D.1) for a

---

15. We assume in this section that our direct utility functions $u_i(\cdot)$ are concave. This is a weak hypothesis (once quasiconcavity has been assumed) which makes sure that in all the optimization problems to be considered, the first-order conditions are sufficient for the determination of global optima. In particular, $v_i(p, \cdot)$ is then a concave function of w_i.

16. In Exercise 4.D.1, you are asked to show that if so desired, problem (4.D.1) can be equivalently formulated as one where social utility is maximized, not by distributing wealth, but by distributing bundles of goods with aggregate value at prices p not larger than w. The fact that in optimally redistributing goods, we can also restrict ourselves to redistributing wealth is, in essence, a version of the second fundamental theorem of welfare economics, which will be covered extensively in Chapter 16.

given value of (p, w). Neglecting boundary solutions, these require that for some $\lambda \geq 0$, we have

$$\lambda = \frac{\partial W}{\partial v_1}\frac{\partial v_1}{\partial w_1} = \cdots = \frac{\partial W}{\partial v_I}\frac{\partial v_I}{\partial w_I} \tag{4.D.2}$$

(For notational convenience, we have omitted the points at which the derivatives are evaluated.) Condition (4.D.2) simply says that at a socially optimal wealth distribution, the social utility of an extra unit of wealth is the same irrespective of who gets it.

By Roy's identity, we have $x_R(p, w) = -[1/(\partial v(p, w)/\partial w)]\,\nabla_p v(p, w)$. Since $v(p, w)$ is the value function of problem (4.D.1), we know that $\partial v/\partial w = \lambda$. (See Section M.K of the Mathematical Appendix) In addition, for any commodity ℓ, the chain rule and (4.D.2)—or, equivalently, the envelope theorem—give us

$$\frac{\partial v}{\partial p_\ell} = \sum_i \frac{\partial W}{\partial v_i}\frac{\partial v_i}{\partial p_\ell} + \lambda \sum_i \frac{\partial w_i}{\partial p_\ell} = \sum_i \frac{\partial W}{\partial v_i}\frac{\partial v_i}{\partial p_\ell},$$

where the second equality follows because $\sum_i w_i(p, w) = w$ for all (p, w) implies that $\sum_i(\partial w_i/\partial p_\ell) = 0$. Hence, in matrix notation, we have

$$\nabla_p v(p, w) = \textstyle\sum_i(\partial W/\partial v_i)\,\nabla_p v_i(p, w_i(p, w)).$$

Finally, using Roy's identity and the first-order condition (4.D.2), we get

$$\begin{aligned} x_R(p, w) &= -\frac{1}{\lambda}\textstyle\sum_i\left[\dfrac{\lambda}{\partial v_i/\partial w_i}\right]\nabla_p v_i(p, w_i(p, w)) \\ &= -\textstyle\sum_i\left[\dfrac{1}{\partial v_i/\partial w_i}\right]\nabla_p v_i(p, w_i(p, w)) \\ &= \textstyle\sum_i x_i(p, w_i(p, w)) = x(p, w), \end{aligned}$$

as we wanted to show. ■

Equipped with Proposition 4.D.1, we can now define a *normative representative consumer.*

Definition 4.D.3: The positive representative consumer $\succsim$ for the aggregate demand $x(p, w) = \sum_i x_i(p, w_i(p, w))$ is a *normative representative consumer* relative to the social welfare function $W(\cdot)$ if for every (p, w), the distribution of wealth $(w_1(p, w), \ldots, w_I(p, w))$ solves problems (4.D.1) and, therefore, the value function of problem (4.D.1) is an indirect utility function for $\succsim$.

If there is a normative representative consumer, the preferences of this consumer have welfare significance and the aggregate demand function $x(p, w)$ can be used to make welfare judgments by means of the techniques described in Section 3.I. In doing so, however, it should never be forgotten that a given wealth distribution rule [the one that solves (4.D.1) for the given social welfare function] is being adhered to and that the "level of wealth" should always be understood as the "optimally distributed level of wealth." For further discussion, see Samuelson (1956) and Chipman and Moore (1979).

Example 4.D.1: Suppose that consumers all have homothetic preferences represented by utility functions homogeneous of degree one. Consider now the social welfare function $W(u_1, \ldots, u_I) = \sum_i \alpha_i \ln u_i$ with $\alpha_i > 0$ and $\sum_i \alpha_i = 1$. Then the optimal

wealth distribution function [for problem (4.D.1)] is the price-independent rule that we adopted in Section 4.C: $w_i(p, w) = \alpha_i w$. (You are asked to demonstrate this fact in Exercise 4.D.6.) Therefore, in the homothetic case, the aggregate demand $x(p, w) = \sum_i x_i(p, \alpha_i w)$ can be viewed as originating from the normative representative consumer generated by this social welfare function. ■

Example 4.D.2: Suppose that all consumers' preferences have indirect utilities of the Gorman form $v_i(p, w_i) = a_i(p) + b(p)w_i$. Note that $b(p)$ does not depend on i, and recall that this includes as a particular case the situation in which preferences are quasilinear with respect to a common numeraire. From Section 4.B, we also know that aggregate demand $x(p, w)$ is independent of the distribution of wealth.[17]

Consider now the *utilitarian* social welfare function $\sum_i u_i$. Then *any* wealth distribution rule $(w_1(p, w), \ldots, w_I(p, w))$ solves the optimization problem (4.D.1), and the indirect utility function that this problem generates is simply $v(p, w) = \sum_i a_i(p) + b(p)w$. (You are asked to show these facts in Exercise 4.D.7.) One conclusion is, therefore, that when indirect utility functions have the Gorman form [with common $b(p)$] and the social welfare function is utilitarian, then aggregate demand can *always* by viewed as being generated by a normative representative consumer.

When consumers have Gorman-form indirect utility functions [with common $b(p)$], the theory of the normative representative consumer admits an important strengthening. In general, the preferences of the representative consumer depend on the form of the social welfare function. *But not in this case.* We now verify that if the indirect utility functions of the consumers have the Gorman form [with common $b(p)$], then the preferences of the representative consumer are independent of the particular social welfare function used.[18] In fact, we show that $v(p, w) = \sum_i a_i(p) + b(p)w$ is an admissible indirect utility function for the normative representative consumer relative to *any* social welfare function $W(u_1, \ldots, u_I)$.

To verify this claim, consider a particular social welfare function $W(\cdot)$, and denote the value function of problem (4.D.1), relative to $W(\cdot)$, by $v^*(p, w)$. We must show that the ordering induced by $v(\cdot)$ and $v^*(\cdot)$ is the same, that is, that for any pair (p, w) and (p', w') with $v(p, w) < v(p', w')$, we have $v^*(p, w) < v^*(p', w')$. Take the vectors of individual wealths $(w_1, \ldots, w_I)$ and $(w'_1, \ldots, w'_I)$ reached as optima of (4.D.1), relative to $W(\cdot)$, for (p, w) and (p', w'), respectively. Denote $u_i = a_i(p) + b(p)w_i$, $u'_i = a_i(p) + b(p)w'_i$, $u = (u_1, \ldots, u_I)$, and $u' = (u'_1, \ldots, u'_I)$. Then $v^*(p, w) = W(u)$ and $v^*(p', w') = W(u')$. Also $v(p, w) = \sum_i a_i(p) + b(p)w = \sum_i u_i$, and similarly, $v(p', w') = \sum_i u'_i$. Therefore, $v(p, w) < v(p', w')$ implies $\sum_i u_i < \sum_i u'_i$. We argue that $\nabla W(u') \cdot (u - u') < 0$, which, $W(\cdot)$ being concave, implies the desired result, namely $W(u) < W(u')$.[19] By expression (4.D.2), at an optimum we have $(\partial W/\partial v_i)(\partial v_i/\partial w_i) = \lambda$ for all i. But in our case, $\partial v_i/\partial w_i = b(p)$ for all i. Therefore, $\partial W/\partial v_i = \partial W/\partial v_j > 0$ for any i, j. Hence, $\sum_i u_i < \sum_i u'_i$ implies $\nabla W(u') \cdot (u - u') < 0$.

The previous point can perhaps be better understood if we observe that when

17. As usual, we neglect the nonnegativity constraints on consumption.

18. But, of course, the optimal distribution rules will typically depend on the social welfare function. Only for the utilitarian social welfare function will it not matter how wealth is distributed.

19. Indeed, concavity of $W(\cdot)$ implies $W(u') + \nabla W(u') \cdot (u - u') \geq W(u)$; see Section M.C of the Mathematical Appendix.

preferences have the Gorman form [with common $b(p)$], then (p', w') is socially better than (p, w) for the utilitarian social welfare function $\sum_i u_i$ *if and only if* when compared with (p, w), (p', w') passes the following *potential compensation test*: For any distribution $(w_1, \ldots, w_I)$ of w, there is a distribution $(w'_1, \ldots, w'_I)$ of w' such that $v_i(p', w'_i) > v_i(p, w_i)$ for all i. To verify this is straightforward. Suppose that

$$(\textstyle\sum_i a_i(p') + b(p')w') - (\sum_i a_i(p) + b(p)w) = c > 0.$$

Then the wealth levels w'_i implicitly defined by $a_i(p') + b(p')w'_i = a_i(p) + b(p)w_i + c/I$ will be as desired.[20] Once we know that (p', w') when compared with (p, w) passes the potential compensation test, it follows merely from the definition of the optimization problem (4.D.1) that (p', w') is better than (p, w) for any normative consumer, that is, for any social welfare function that we may wish to employ (see Exercise 4.D.8).

The two properties just presented—independence of the representative consumer's preferences from the social welfare function and the potential compensation criterion—will be discussed further in Sections 10.F and 22.C. For the moment, we simply emphasize that they are not general properties of normative representative consumers. By choosing the distribution rules that solve (4.D.1), we can generate a normative representative consumer for any set of individual utilities and any social welfare function. For the properties just reviewed to hold, the individual preferences have been required to have the Gorman form [with common $b(p)$]. ■

It is important to stress the distinction between the concepts of a positive and a normative representative consumer. It is *not* true that whenever aggregate demand can be generated by a positive representative consumer, this representative consumer's preferences have normative content. It may even be the case that a positive representative consumer exists but that there is *no* social welfare function that leads to a normative representative consumer. We expand on this point in the next few paragraphs [see also Dow and Werlang (1988) and Jerison (1994)].

We are given a distribution rule $(w_1(p, w), \ldots, w_I(p, w))$ and assume that a positive representative consumer with utility function $u(x)$ exists for the aggregate demand $x(p, w) = \sum_i x_i(p, w_i(p, w))$. In principle, using the integrability techniques presented in Section 3.H, it should be possible to determine the preferences of the representative consumer from the knowledge of $x(p, w)$. Now fix any $(\bar{p}, \bar{w})$, and let $\bar{x} = x(\bar{p}, \bar{w})$. Relative to the aggregate consumption vector $\bar{x}$, we can define an at-least-as-good-as set for the representative consumer:

$$B = \{x \in \mathbb{R}^L_+ : u(x) \geq (\bar{x})\} \subset \mathbb{R}^L_+.$$

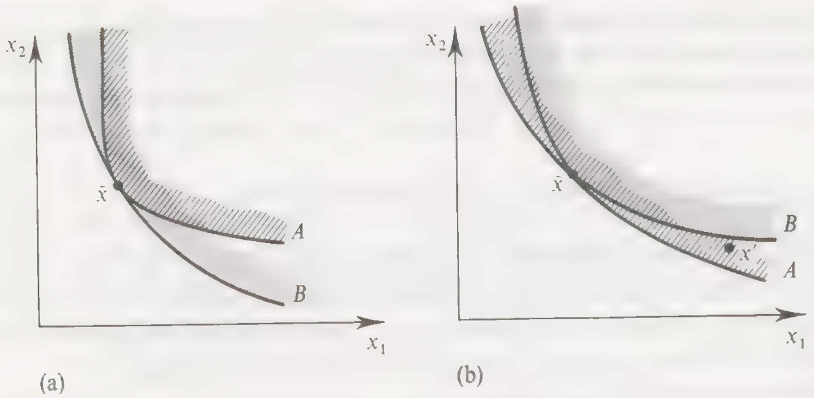
Next, let $\bar{w}_i = w_i(\bar{p}, \bar{w})$ and $\bar{x}_i = x_i(\bar{p}, \bar{w}_i)$, and consider the set

$$A = \{x = \textstyle\sum_i x_i : x_i \succsim_i \bar{x}_i \text{ for all } i\} \subset \mathbb{R}^L_+.$$

In words, A is the set of aggregate consumption vectors for which there is a distribution of commodities among consumers that makes every consumer as well off as under $(\bar{x}_1, \ldots, \bar{x}_I)$. The boundary of this set is sometimes called a *Scitovsky contour*. Note that both set A and set B are supported by the price vector $\bar{p}$ at $\bar{x}$ (see Figure 4.D.1).

If the given wealth distribution comes from the solution to a social welfare optimization problem of the type (4.D.1) (i.e., if the positive representative consumer is in fact a normative

20. We continue to neglect nonnegativity constraints on wealth.



$A = \{x \in \Sigma_i\, x_i\colon u_i(x_i) \geq u_i(\bar{x}_i) \text{ for all } i\}$
$B = \{x \in \mathbb{R}^2_+\colon u(x) \geq u(\bar{x})\}$

Figure 4.D.1 Comparing the at-least-as-good-as set of the positive representative consumer with the sum of the at-least-as-good-as sets of the individual consumers. (a) The positive representative consumer could be a normative representative consumer. (b) The positive representative consumer cannot be a normative representative consumer.

representative consumer), then this places an important restriction on how sets A and B relate to each other: Every element of set A must be an element of set B. This is so because the social welfare function underlying the normative representative consumer is increasing in the utility level of every consumer (and thus any aggregate consumption bundle that could be distributed in a manner that guarantees to every consumer a level of utility as high as the levels corresponding to the optimal distribution of $\bar{x}$ must receive a social utility higher than the latter; see Exercise 4.D.4). That is, a *necessary* condition for the existence of a normative representative consumer is that $A \subset B$. A case that satisfies this necessary condition is depicted in Figure 4.D.1(a).

However, there is nothing to prevent the existence, in a particular setting, of a positive representative consumer with a utility function $u(x)$ that fails to satisfy this condition, as in Figure 4.D.1(b). To provide some further understanding of this point, Exercise 4.D.9 asks you to show that $A \subset B$ implies that $\sum_i S_i(\bar{p}, \bar{w}_i) - S(\bar{p}, \bar{w})$ is positive semidefinite, where $S(p, w)$ and $S_i(p, w_i)$ are the Slutsky matrices of aggregate and individual demand, respectively. Informally, we could say that the substitution effects of aggregate demand must be larger in absolute value than the sum of individual substitution effects (geometrically, this corresponds to the boundary of B being flatter at $\bar{x}$ than the boundary of A). This observation allows us to generate in a simple manner examples in which aggregate demand can be rationalized by preferences but, nonetheless, there is no normative representative consumer.

Suppose, for example, that the wealth distribution rule is of the form $w_i(p, w) = \alpha_i w$. Suppose also that $S(p, w)$ happens to be symmetric for all (p, w); if $L = 2$, this is automatically satisfied. Then, from integrability theory (see Section 3.H), we know that a sufficient condition for the existence of underlying preferences is that, for all (p, w), we have $dp \cdot S(p, w)\, dp < 0$ for all $dp \neq 0$ not proportional to p (we abbreviate this as the *n.d. property*). On the other hand, as we have just seen, a necessary condition for the existence of a normative representative consumer is that $C(\bar{p}, \bar{w}) = \sum_i S_i(\bar{p}, \bar{w}_i) - S(\bar{p}, \bar{w})$ be positive semidefinite [this is the same matrix discussed in Section 4.C; see expression (4.C.8)]. Thus, if $S(p, w)$ has the n.d. property for all (p, w) but $C(\bar{p}, \bar{w})$ is not positive semidefinite [i.e., wealth effects are such that $S(\bar{p}, \bar{w})$ is "less negative" than $\sum_i S_i(\bar{p}, \bar{w})$], then a positive representative consumer exists that, nonetheless, cannot be made normative for any social welfare function. (Exercise 4.D.10 provides an instance where this is indeed the case.) In any example of this nature we have moves in aggregate consumption that would pass a potential compensation test (each consumer's welfare could be made better off by an appropriate distribution of the move) but are regarded as socially inferior under the utility function that rationalizes aggregate demand. [In Figure 4.D.1(b), this could be the move from $\bar{x}$ to x'.]

The moral of all this is clear: The existence of preferences that explain behavior is not

enough to attach to them any welfare significance. For the latter, it is also necessary that these preferences exist for the right reasons. ■

APPENDIX A: REGULARIZING EFFECTS OF AGGREGATION

This appendix is devoted to making the point that although aggregation can be deleterious to the preservation of the good properties of individual demand, it can also have helpful *regularizing* effects. By regularizing, we mean that the average (per-consumer) demand will tend to be more continuous or smooth, as a function of prices, than the individual components of the sum.

Recall that if preferences are strictly convex, individual demand functions are continuous. As we noted, aggregate demand will then be continuous as well. But average demand can be (nearly) continuous even when individual demands are not. The key requirement is one of *dispersion* of individual preferences.

Example 4.AA.1: Suppose that there are two commodities. Consumers have quasilinear preferences with the second good as numeraire. The first good, on the other hand, is available only in integer amounts, and consumers have no wish for more than one unit of it. Thus, normalizing the utility of zero units of the first good to be zero, the preferences of consumer i are completely described by a number v_{1i}, the utility in terms of numeraire of holding one unit of the first good. It is then clear that the demand for the first good by consumer i is given by the correspondence

$$\begin{aligned} x_{1i}(p_1) &= 1 && \text{if } p_1 < v_{1i}, \\ &= \{0, 1\} && \text{if } p_1 = v_{1i}, \\ &= 0 && \text{if } p_1 > v_{1i}, \end{aligned}$$

which is depicted in Figure 4.AA.1(a). Thus, individual demand exhibits a sudden, discontinuous jump in demand from 0 to 1 as the price crosses the value $p_1 = v_{1i}$.

Suppose now that there are many consumers. In fact, consider the limit situation where there is an actual continuum of consumers. We could then say that individual preferences are *dispersed* if there is no concentrated group of consumers having any particular value of v_1 or, more precisely, if the statistical distribution function of the v_1's, $G(v_1)$, is continuous. Then, denoting by $x_1(p_1)$ the average demand for the first good, we have $x_1(p_1) =$ "mass of consumers with $v_1 > p_1$" $= 1 - G(p_1)$.

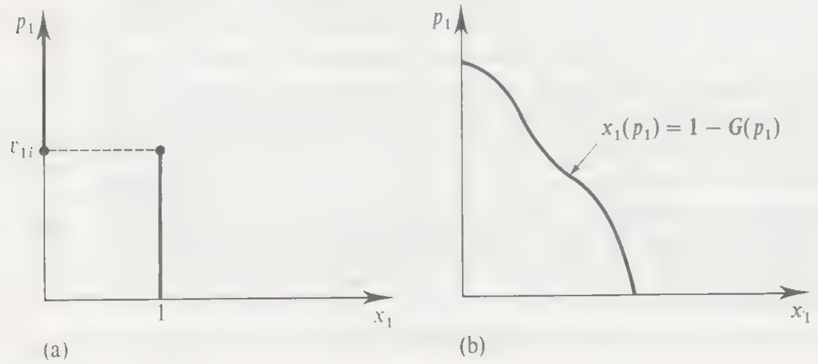


Figure 4.AA.1 The regularizing effect of aggregation. (a) Individual demand. (b) Aggregate demand when the distribution of the v_1's is $G(\cdot)$.

Hence, the aggregate demand $x_1(\cdot)$, shown in Figure 4.AA.1(b), is a nice continuous function even though none of the individual demand correspondences are so. Note that with only a finite number of consumers, the distribution function $G(\cdot)$ cannot quite be a continuous function; but if the consumers are many, then it can be nearly continuous. ∎

The regularizing effects of aggregation are studied again in Section 17.I. We show there that in general (i.e., without dispersedness requirements), the aggregation of numerous individual demand correspondences will generate a (nearly) *convex-valued* average demand correspondence.

REFERENCES

Chipman, J. S., and J. Moore. (1979). On social welfare functions and the aggregation of preferences. *Journal of Economic Theory* **21**: 111–39.

Deaton, A., and J. Muellbauer. (1980). *Economics and Consumer Behavior*. Cambridge, UK: Cambridge University Press.

Dow, J., and S. Werlang. (1988). The consistency of welfare judgments with a representative consumer. *Journal of Economic Theory* **44**: 265–80.

Freixas, X., and A. Mas-Colell. (1987). Engel curves leading to the weak axiom in the aggregate. *Econometrica* **21**: 63–80.

Hildenbrand, W. (1983). On the "law of demand." *Econometrica* **51**: 997–1020.

Hildenbrand, W. (1994). *Market Demand: Theory and Empirical Evidence*. Princeton: Princeton University Press.

Jerison, M. (1982). The representative consumer and the weak axiom when the distribution of income is fixed. Working paper, Department of Economics, SUNY Albany.

Jerison, M. (1994). Optimal income distribution rules and representative consumers. *Review of Economic Studies* **61**: 739–71.

Jorgenson, D. (1990). Aggregate consumer behavior and the measurement of social welfare *Econometrica* **58**: 1007–1030.

Lau, L. (1982). A note on the fundamental theorem of exact aggregation, *Economic letters* **9**: 119–26.

Mas-Colell, A. (1991). On the uniqueness of equilibrium once again. Chap. 12 in *Equilibrium Theory and Applications*, edited by W. Barnett, B. Cornet, C. d'Aspremont, J. Gabszewicz, A. Mas-Colell. Cambridge, U.K.: Cambridge University Press.

Milleron, J. C. (1974). Unicité et stabilité de l'équilibre en économie de distribution. Seminaire d'Econométrie Roy-Malinvaud, preprint.

Mitiushin, L. G., and W. M. Polterovich (1978). Criteria for monotonicity of demand functions [in Russian]. *Ekonomika i Matematichskie Metody* **14**: 122–28.

Samuelson, P. (1956). Social indifference curves. *Quarterly Journal of Economics* **70**: 1–22.

EXERCISES

4.B.1[B] Prove the sufficiency part of Proposition 4.B.1. Show also that if preferences admit Gorman-form indirect utility functions with the same $b(p)$, then preferences admit expenditure functions of the form $e_i(p, u_i) = c(p)u_i + d_i(p)$.

4.B.2[B] Suppose that there are I consumers and L commodities. Consumers differ only by their wealth levels w_i and by a taste parameter s_i, which we could call *family size*. Thus, denote the indirect utility function of consumer i by $v(p, w_i, s_i)$. The corresponding Walrasian demand function for consumer i is $x(p, w_i, s_i)$.

(a) Fix $(s_1, \ldots, s_I)$. Show that if for any $(w_1, \ldots, w_I)$ aggregate demand can be written as a function of only p and aggregate wealth $w = \sum_i w_i$ (or, equivalently, average wealth), and if every consumer's preference relationship $\succsim_i$ is homothetic, then all these preferences must be identical [and so $x(p, w_i, s_i)$ must be independent of s_i].

(b) Give a sufficient condition for aggregate demand to depend only on aggregate wealth w and $\sum_i s_i$ (or, equivalently, average wealth and average family size).

4.C.1$^{\mathrm{C}}$ Prove that if $x_i(p, w_i)$ satisfies the ULD, then $D_p x_i(p, w_i)$ is negative semidefinite [i.e., $dp \cdot D_p x_i(p, w_i)\, dp \leq 0$ for all dp]. Also show that if $D_p x_i(p, w_i)$ is negative definite for all p, then $x_i(p, w_i)$ satisfies the ULD (this second part is harder).

4.C.2$^{\mathrm{A}}$ Prove a version of Proposition 4.C.1 by using the (sufficient) differential versions of the ULD and the WA. (Recall from the small type part of Section 2.F that a sufficient condition for the WA is that $v \cdot S(p, w) v < 0$ whenever v is not proportional to p.)

4.C.3$^{\mathrm{A}}$ Give a graphical two-commodity example of a preference relation generating a Walrasian demand that does not satisfy the ULD property. Interpret.

4.C.4$^{\mathrm{C}}$ Show that if the preference relation $\succsim_i$ on $\mathbb{R}^2_+$ has L-shaped indifference curves and the demand function $x_i(p, w_i)$ has the ULD property, then $\succsim_i$ must be homothetic. [*Hint*: The L shape of indifference curves implies $S_i(p, w_i) = 0$ for all (p, w_i); show that if $D_{w_i} x_i(\bar{p}, \bar{w}_i) \neq (1/\bar{w}_i) x_i(\bar{p}, \bar{w}_i)$, then there is $v \in \mathbb{R}^L$ such that $v \cdot D_p x_i(\bar{p}, \bar{w}_i) v > 0$.]

4.C.5$^{\mathrm{C}}$ Prove Proposition 4.C.3. To that effect, you can fix $w = 1$. The proof is best done in terms of the indirect demand function $g_i(x) = (1/x \cdot \nabla u_i(x))\, \nabla u_i(x)$ [note that $x = x_i(g_i(x), 1)$]. For an individual consumer, the ULD is self-dual; that is, it is equivalent to $(g_i(x) - g_i(y)) \cdot (x - y) < 0$ for all $x \neq y$. In turn, this property is implied by the negative definiteness of $Dg_i(x)$ for all x. Hence, concentrate on proving this last property. More specifically, let $v \neq 0$, and denote $q = \nabla u_i(x)$ and $C = D^2 u_i(x)$. You want to prove $v \cdot Dg_i(x) v < 0$. [*Hint*: You can first assume $q \cdot v = q \cdot x$; then differentiate $g_i(x)$, and make use of the equality $v \cdot Cv - x \cdot Cv = (v - \frac{1}{2}x) \cdot C(v - \frac{1}{2}x) - \frac{1}{4} x \cdot Cx$.]

4.C.6$^{\mathrm{A}}$ Show that if $u_i(x_i)$ is homogeneous of degree one, so that $\succsim_i$ is homothetic, then $\sigma_i(x_i) = 0$ for all x_i [$\sigma_i(x_i)$ is the quotient defined in Proposition 4.C.3].

4.C.7$^{\mathrm{B}}$ Show that Proposition 4.C.4 still holds if the distribution of wealth has a nonincreasing density function on $[0, \bar{w}]$. A more realistic distribution of wealth would be *unimodal* (i.e., an increasing and then decreasing density function with a single peak). Argue that there are unimodal distributions for which the conclusions of the proposition do not hold.

4.C.8$^{\mathrm{A}}$ Derive expression (4.C.7), the aggregate version of the Slutsky matrix.

4.C.9$^{\mathrm{A}}$ Verify that if individual preferences $\succsim_i$ are homothetic, then the matrix $C(p, w)$ defined in expression (4.C.8) is positive semidefinite.

4.C.10$^{\mathrm{C}}$ Argue that for the Hildenbrand example studied in Proposition 4.C.4, $C(p, w)$ is positive semidefinite. Conclude that aggregate demand satisfies the WA for that wealth distribution. [Note: You must first adapt the definition of $C(p, w)$ to the continuum-of-consumers situation of the example.]

4.C.11$^{\mathrm{B}}$ Suppose there are two consumers, 1 and 2, with utility functions over two goods, 1 and 2, of $u_1(x_{11}, x_{21}) = x_{11} + 4\sqrt{x_{21}}$ and $u_2(x_{12}, x_{22}) = 4\sqrt{x_{12}} + x_{22}$. The two consumers have identical wealth levels $w_1 = w_2 = w/2$.

(a) Calculate the individual demand functions and the aggregate demand function.

(b) Compute the individual Slutsky matrices $S_i(p, w/2)$ (for $i = 1, 2$) and the aggregate

Slutsky matrix $S(p, w)$. [*Hint*: Note that for this two-good example, only one element of each matrix must be computed to determine the entire matrix.] Show that $dp \cdot S(p, w)\, dp < 0$ for all $dp \neq 0$ not proportional to p. Conclude that aggregate demand satisfies the *WA*.

(c) Compute the matrix $C(p, w) = \sum_i S_i(p, w/2) - S(p, w)$ for prices $p_1 = p_2 = 1$. Show that it is positive semidefinite if $w > 16$ and that it is negative semidefinite if $8 < w < 16$. In fact, argue that in the latter case, $dp \cdot C(p, w)\, dp < 0$ for some dp [so that $C(p, w)$ is *not* positive semidefinite]. Conclude that $C(p, w)$ positive semidefinite is not necessary for the WA to be satisfied.

(d) For each of the two cases $w > 16$ and $8 < w < 16$, draw a picture in the (x_1, x_2) plane depicting each consumer's consumption bundle and his wealth expansion path for the prices $p_1 = p_2 = 1$. Compare your picture with Figure 4.C.2.

4.C.12[B] The results presented in Sections 4.B and 4.C indicate that if for any $(w_1, \ldots, w_I)$ aggregate demand can be written as a function of only aggregate wealth [i.e., as $x(p, \sum_i w_i)$], then aggregate demand must satisfy the WA. The *distribution function* $F\colon [0, \infty) \to [0, 1]$ of $(w_1, \ldots, w_I)$ is defined as $F(w) = (1/I)$(*number of* i's with $w_i \leq w$) for any w. Suppose now that for any $(w_1, \ldots, w_I)$, aggregate demand can be written as a function of the corresponding aggregate *distribution* $F(\cdot)$ of wealth. Show that aggregate demand does not necessarily satisfy the WA. [*Hint*: It suffices to give a two-commodity, two-consumer example where preferences are identical, wealths are $w_1 = 1$ and $w_2 = 3$, and the WA fails. Try to construct the example graphically. It is a matter of making sure that four suitably positioned indifference curves can be fitted together without crossing.]

4.C.13[C] Consider a two-good environment with two consumers. Let the wealth distribution rule be $w_1(p, w) = wp_1/(p_1 + p_2)$, $w_2(p, w) = wp_2/(p_1 + p_2)$. Exhibit an example in which the two consumers have homothetic preferences but, nonetheless, the aggregate demand fails to satisfy the weak axiom. A good picture will suffice. Why does not Proposition 4.C.1 apply?

4.D.1[B] In this question we are concerned with a normative representative consumer. Denote by $v(p, w)$ the optimal value of problem (4.D.1), and by $(w_1(p, w), \ldots, w_I(p, w))$ the corresponding optimal wealth distribution rules. Verify that $v(p, w)$ is also the optimal value of

$$\begin{aligned} \underset{x_1, \ldots, x_I}{\text{Max}} \quad & W(u_1(x_1), \ldots, u_I(x_I)) \\ \text{s.t. } & p \cdot (\textstyle\sum_i x_i) \leq w \end{aligned}$$

and that $[x_1(p, w_1(p, w)), \ldots, x_I(p, w_I(p, w))]$ is a solution to this latter problem. Note the implication that to maximize social welfare given prices p and wealth w, the planner need not control consumption directly, but rather need only distribute wealth optimally and allow consumers to make consumption decisions independently given prices p.

4.D.2[B] Verify that $v(p, w)$, defined as the optimal value of problem (4.D.1), has the properties of an indirect utility function (i.e., that it is homogeneous of degree zero, increasing in w and decreasing in p, and quasiconvex).

4.D.3[B] It is good to train one's hand in the use of inequalities and the Kuhn–Tucker conditions. Prove Proposition 4.D.1 again, this time allowing for corner solutions.

4.D.4[C] Suppose that there is a normative representative consumer with wealth distribution rule $(w_1(p, w), \ldots, w_I(p, w))$. For any $x \in \mathbb{R}^L_+$, define

$$\begin{aligned} u(x) = \underset{(x_1, \ldots, x_I)}{\text{Max}} \quad & W(u_1(x_1), \ldots, u_I(x_I)) \\ \text{s.t. } & \textstyle\sum_i x_i \leq x. \end{aligned}$$

(a) Give conditions implying that $u(\cdot)$ has the properties of a utility function; that is, it is monotone, continuous, and quasiconcave (and even concave).

(b) Show that for any (p, w), the Walrasian demand generated from the problem $\text{Max}_x\, u(x)$ s.t. $p \cdot x \leq w$ is equal to the aggregate demand function.

4.D.5^A Suppose that there are I consumers and that consumer i's utility function is $u_i(x_i)$, with demand function $x_i(p, w_i)$. Consumer i's wealth w_i is generated according to a wealth distribution rule $w_i = \alpha_i w$, where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Provide an example (i.e., a set of utility functions) in which this economy does *not* admit a positive representative consumer.

4.D.6^B Establish the claims made in Example 4.D.1.

4.D.7^B Establish the claims made in the second paragraph of Example 4.D.2.

4.D.8^A Say that (p', w') passes the *potential compensation test* over (p, w) if for any distribution $(w_1, \ldots, w_I)$ of w there is a distribution $(w'_1, \ldots, w'_I)$ of w' such that $v_i(p', w'_i) > v_i(p, w_i)$ for all i. Show that if $(p'w')$ passes the potential compensation test over (p, w), any normative representative consumer must prefer (p', w') over (p, w).

4.D.9^B Show that $A \subset B$ (notation as in Section 4.D) implies that $S(\bar{p}, \bar{w}) - \sum_i S_i(\bar{p}, \bar{w}_i)$ is negative semidefinite. [*Hint*: Consider $g(p) = e(p, u(\bar{x})) - \sum_i e_i(p, u_i(\bar{x}_i))$, where $e(\cdot)$ is the expenditure function for $u(\cdot)$ and $e_i(\cdot)$ is the expenditure function for $u_i(\cdot)$. Note that $A = \sum_i \{x_i : u_i(x_i) \geq u_i(\bar{x}_i)\}$ implies that $\sum_i e_i(p, u_i(\bar{x}_i))$ is the optimal value of the problem $\text{Min}_{x \in A}\, p \cdot x$. From this and $A \subset B$, you get $g(p) \leq 0$ for all p and $g(\bar{p}) = 0$. Therefore, $D^2 g(\bar{p})$ is negative semidefinite. Show then that $D^2 g(\bar{p}) = S(\bar{p}, \bar{w}) - \sum_i S_i(\bar{p}, \bar{w}_i)$.]

4.D.10^A Argue that in the example considered in Exercise 4.C.11, there is a positive representative consumer rationalizing aggregate demand but that there cannot be a normative representative consumer.

4.D.11^C Argue that for $L > 2$, the Hildenbrand case of Proposition 4.C.4 need not admit a positive representative consumer. [*Hint*: Argue that the Slutsky matrix may fail to be symmetric.]

CHAPTER

5

Production

5.A Introduction

In this chapter, we move to the supply side of the economy, studying the process by which the goods and services consumed by individuals are produced. We view the supply side as composed of a number of productive units, or, as we shall call them, "firms." Firms may be corporations or other legally recognized businesses. But they must also represent the productive possibilities of individuals or households. Moreover, the set of all firms may include some potential productive units that are never actually organized. Thus, the theory will be able to accommodate both active production processes and potential but inactive ones.

Many aspects enter a full description of a firm: Who owns it? Who manages it? How is it managed? How is it organized? What can it do? Of all these questions, we concentrate on the last one. Our justification is not that the other questions are not interesting (indeed, they are), but that we want to arrive as quickly as possible at a minimal conceptual apparatus that allows us to analyze market behavior. Thus, our model of production possibilities is going to be very parsimonious: The firm is viewed merely as a "black box", able to transform inputs into outputs.

In Section 5.B, we begin by introducing the firm's *production set*, a set that represents the production *activities*, or *production plans*, that are technologically feasible for the firm. We then enumerate and discuss some commonly assumed properties of production sets, introducing concepts such as *returns to scale*, *free disposal*, and *free entry*.

After studying the firm's technological possibilities in Section 5.B, we introduce its objective, the goal of *profit maximization*, in Section 5.C. We then formulate and study the firm's profit maximization problem and two associated objects, the firm's *profit function* and its *supply correspondence*. These are, respectively, the value function and the optimizing vectors of the firm's profit maximization problem. Related to the firm's goal of profit maximization is the task of achieving cost-minimizing production. We also study the firm's cost minimization problem and two objects associated with it: The firm's *cost function* and its *conditional factor demand correspondence*. As with the utility maximization and expenditure minimization problems in the theory of demand, there is a rich duality theory associated with the profit maximization and cost minimization problems.

Section 5.D analyzes in detail the geometry associated with cost and production relationships for the special but theoretically important case of a technology that produces a single output.

Aggregation theory is studied in Section 5.E. We show that aggregation on the supply side is simpler and more powerful than the corresponding theory for demand covered in Chapter 4.

Section 5.F constitutes an excursion into welfare economics. We define the concept of *efficient production* and study its relation to profit maximization. With some minor qualifications, we see that profit-maximizing production plans are efficient and that when suitable convexity properties hold, the converse is also true: An efficient plan is profit maximizing for an appropriately chosen vector of prices. This constitutes our first look at the important ideas of the *fundamental theorems of welfare economics.*

In Section 5.G, we point out that profit maximization does not have the same primitive status as preference maximization. Rigorously, it should be derived from the latter. We discuss this point and related issues.

In Appendix A, we study in more detail a particular, important case of production technologies: Those describable by means of linear constraints. It is known as the *linear activity model.*

5.B Production Sets

As in the previous chapters, we consider an economy with L commodities. A *production vector* (also known as an *input–output*, or *netput*, vector, or as a *production plan*) is a vector $y = (y_1, \dots, y_L) \in \mathbb{R}^L$ that describes the (net) outputs of the L commodities from a production process. We adopt the convention that positive numbers denote outputs and negative numbers denote inputs. Some elements of a production vector may be zero; this just means that the process has no net output of that commodity.

Example 5.B.1: Suppose that $L = 5$. Then $y = (-5, 2, -6, 3, 0)$ means that 2 and 3 units of goods 2 and 4, respectively, are produced, while 5 and 6 units of goods 1 and 3, respectively, are used. Good 5 is neither produced nor used as an input in this production vector. ■

To analyze the behavior of the firm, we need to start by identifying those production vectors that are technologically possible. The set of all production vectors that constitute feasible plans for the firm is known as the *production set* and is denoted by $Y \subset \mathbb{R}^L$. Any $y \in Y$ is possible; any $y \notin Y$ is not. The production set is taken as a primitive datum of the theory.

The set of feasible production plans is limited first and foremost by technological constraints. However, in any particular model, legal restrictions or prior contractual commitments may also contribute to the determination of the production set.

It is sometimes convenient to describe the production set Y using a function $F(\cdot)$, called the *transformation function*. The transformation function $F(\cdot)$ has the property that $Y = \{y \in \mathbb{R}^L\colon F(y) \leq 0\}$ and $F(y) = 0$ if and only if y is an element of the boundary of Y. The set of boundary points of Y, $\{y \in \mathbb{R}^L\colon F(y) = 0\}$, is known as the *transformation frontier*. Figure 5.B.1 presents a two-good example.

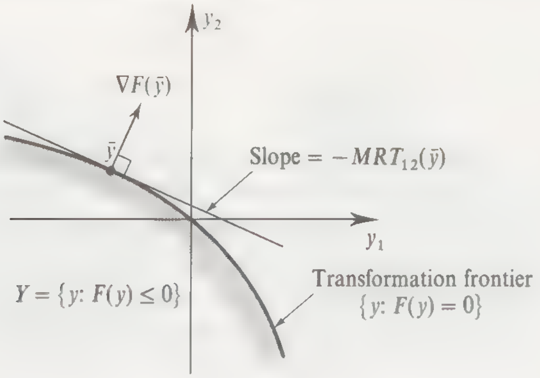


Figure 5.B.1
The production set and transformation frontier.

If $F(\cdot)$ is differentiable, and if the production vector $\bar{y}$ satisfies $F(\bar{y}) = 0$, then for any commodities ℓ and k, the ratio

$$MRT_{\ell k}(\bar{y}) = \frac{\partial F(\bar{y})/\partial y_\ell}{\partial F(\bar{y})/\partial y_k}$$

is called the *marginal rate of transformation* (MRT) *of good* ℓ *for good* k *at* $\bar{y}$.[1] The marginal rate of transformation is a measure of how much the (net) output of good k can increase if the firm decreases the (net) output of good ℓ by one marginal unit. Indeed, from $F(\bar{y}) = 0$, we get

$$\frac{\partial F(\bar{y})}{\partial y_k} dy_k + \frac{\partial F(\bar{y})}{\partial y_\ell} dy_\ell = 0,$$

and therefore the slope of the transformation frontier at $\bar{y}$ in Figure 5.B.1 is precisely $-MRT_{12}(\bar{y})$.

Technologies with Distinct Inputs and Outputs

In many actual production processes, the set of goods that can be outputs is distinct from the set that can be inputs. In this case, it is sometimes convenient to notationally distinguish the firm's inputs and outputs. We could, for example, let $q = (q_1, \ldots, q_M) \geq 0$ denote the production levels of the firm's M outputs and $z = (z_1, \ldots, z_{L-M}) \geq 0$ denote the amounts of the firm's $L - M$ inputs, with the convention that the amount of input z_ℓ used is now measured as a *nonnegative* number (as a matter of notation, we count all goods not actually used in the process as inputs).

One of the most frequently encountered production models is that in which there is a single output. A single-output technology is commonly described by means of a *production function* $f(z)$ that gives the maximum amount q of output that can be produced using input amounts $(z_1, \ldots, z_{L-1}) \geq 0$. For example, if the output is good L, then (assuming that output can be disposed of at no cost) the production function $f(\cdot)$ gives rise to the production set:

$$Y = \{(-z_1, \ldots, -z_{L-1}, q): q - f(z_1, \ldots, z_{L-1}) \leq 0 \quad \text{and} \quad (z_1, \ldots, z_{L-1}) \geq 0\}.$$

Holding the level of output fixed, we can define the *marginal rate of technical*

1. As in Chapter 3, in computing ratios such as this, we always assume that $\partial F(\bar{y})/\partial y_k \neq 0$.

substitution (MRTS) of input ℓ *for input k at* $\bar{z}$ as

$$MRTS_{\ell k}(\bar{z}) = \frac{\partial f(\bar{z})/\partial z_\ell}{\partial f(\bar{z})/\partial z_k}.$$

The number $MRTS_{\ell k}(\bar{z})$ measures the additional amount of input k that must be used to keep output at level $\bar{q} = f(\bar{z})$ when the amount of input ℓ is decreased marginally. It is the production theory analog to the consumer's marginal rate of substitution. In consumer theory, we look at the trade-off between commodities that keeps utility constant, here, we examine the trade-off between inputs that keeps the amount of output constant. Note that $MRTS_{\ell k}$ is simply a renaming of the marginal rate of transformation of input ℓ for input k in the special case of a single-output, many-input technology.

Example 5.B.2: *The Cobb–Douglas Production Function* The *Cobb–Douglas* production function with two inputs is given by $f(z_1, z_2) = z_1^\alpha z_2^\beta$, where $\alpha \geq 0$ and $\beta \geq 0$. The marginal rate of technical substitution between the two inputs at $z = (z_1, z_2)$ is $MRTS_{12}(z) = \alpha z_2/\beta z_1$. ■

Properties of Production Sets

We now introduce and discuss a fairly exhaustive list of commonly assumed properties of production sets. The appropriateness of each of these assumptions depends on the particular circumstances (indeed, some of them are mutually exclusive).[2]

(i) *Y is nonempty.* This assumption simply says that the firm has something it can plan to do. Otherwise, there is no need to study the behavior of the firm in question.

(ii) *Y is closed.* The set Y includes its boundary. Thus, the limit of a sequence of technologically feasible input–output vectors is also feasible; in symbols, $y^n \to y$ and $y^n \in Y$ imply $y \in Y$. This condition should be thought of as primarily technical.[3]

(iii) *No free lunch.* Suppose that $y \in Y$ and $y \geq 0$, so that the vector y does not use any inputs. The no-free-lunch property is satisfied if this production vector cannot produce output either. That is, whenever $y \in Y$ and $y \geq 0$, then $y = 0$; it is not possible to produce something from nothing. Geometrically, $Y \cap \mathbb{R}_+^L \subset \{0\}$. For $L = 2$, Figure 5.B.2(a) depicts a set that violates the no-free-lunch property, the set in Figure 5.B.2(b) satisfies it.

(iv) *Possibility of inaction* This property says that $0 \in Y$: Complete shutdown is possible. Both sets in Figure 5.B.2, for example, satisfy this property. The point in time at which production possibilities are being analyzed is often important for the validity of this assumption. If we are contemplating a firm that could access a set of technological possibilities but that has not yet been organized, then inaction is clearly

2. For further discussion of these properties, see Koopmans (1957) and Chapter 3 of Debreu (1959).

3. Nonetheless, we show in Exercise 5.B.4 that there is an important case of economic interest when it raises difficulties.

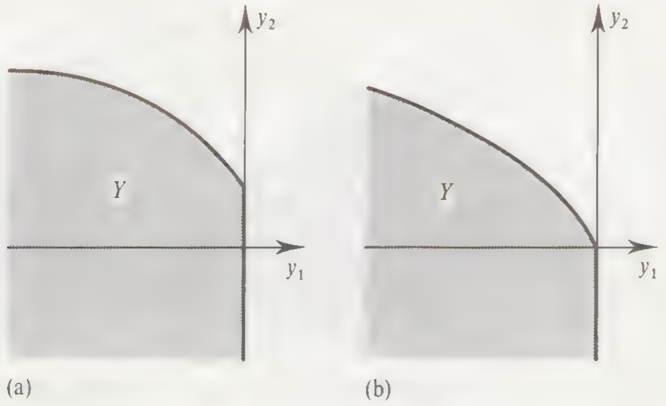


Figure 5.B.2
The no free lunch property.
(a) Violates no free lunch.
(b) Satisfies no free lunch.

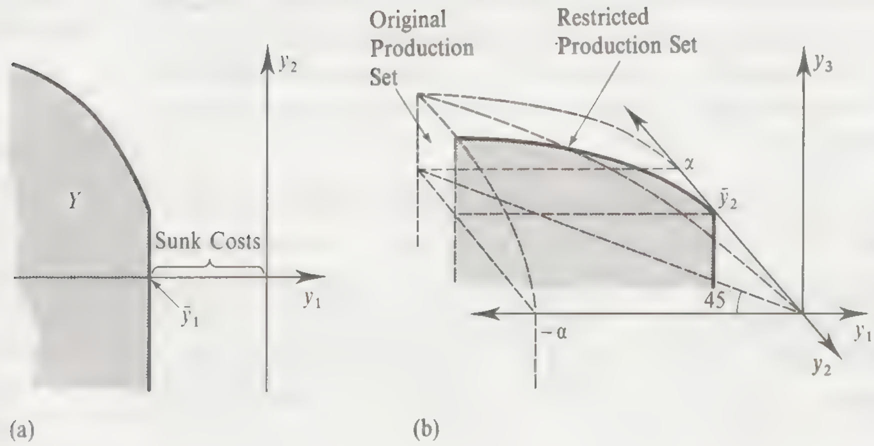
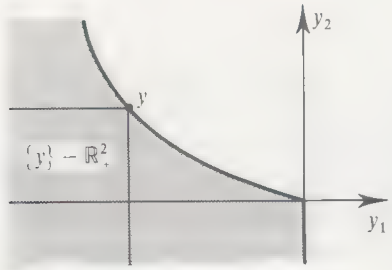


Figure 5.B.3
Two production sets with sunk costs.
(a) A minimal level of expenditure committed.
(b) One kind of input fixed.

possible. But if some production decisions have already been made, or if irrevocable contracts for the delivery of some inputs have been signed, inaction is not possible. In that case, we say that some costs are *sunk*. Figure 5.B.3 depicts two examples. The production set in Figure 5.B.3(a) represents the *interim* production possibilities arising when the firm is already committed to use at least $-\bar{y}_1$ units of good 1 (perhaps because it has already signed a contract for the purchase of this amount); that is, the set is a *restricted production set* that reflects the firm's remaining choices from some original production set Y like the ones in Figure 5.B.2. In Figure 5.B.3(b), we have a second example of sunk costs. For a case with one output (good 3) and two inputs (goods 1 and 2), the figure illustrates the restricted production set arising when the level of the second input has been irrevocably set at $\bar{y}_2 < 0$ [here, in contrast with Figure 5.B.3(a), increases in the use of the input are impossible].

(v) *Free disposal.* The property of free disposal holds if the absorption of any additional amounts of inputs without any reduction in output is always possible. That is, if $y \in Y$ and $y' \leq y$ (so that y' produces at most the same amount of outputs using at least the same amount of inputs), then $y' \in Y$. More succinctly, $Y - \mathbb{R}^L_+ \subset Y$ (see Figure 5.B.4). The interpretation is that the extra amount of inputs (or outputs) can be disposed of or eliminated at no cost.



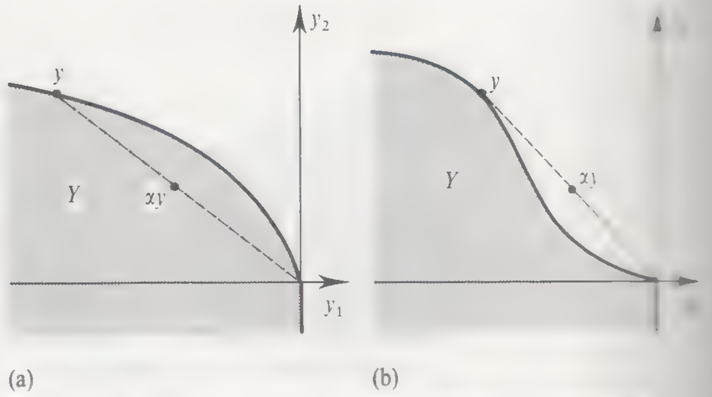


Figure 5.B.4 (left) The free disposal property.

Figure 5.B.5 (right) The nonincreasing returns to scale property. (a) Nonincreasing returns satisfied. (b) Nonincreasing returns violated.

(vi) *Irreversibility*. Suppose that $y \in Y$ and $y \neq 0$. Then irreversiblity says that $-y \notin Y$. In words, it is impossible to reverse a technologically possible production vector to transform an amount of output into the same amount of input that was used to generate it. If, for example, the description of a commodity includes the time of its availability, then irreversibility follows from the requirement that inputs be used before outputs emerge.

Exercise 5.B.1: Draw two production sets: one that violates irreversibility and one that satisfies this property.

decreasing

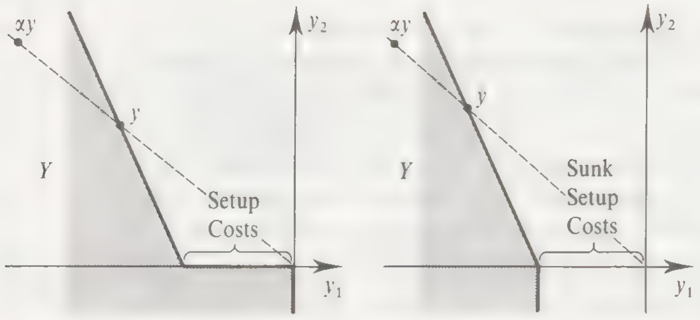
(vii) *Nonincreasing returns to scale*. The production technology Y exhibits nonincreasing returns to scale if for any $y \in Y$, we have $\alpha y \in Y$ for all scalars $\alpha \in [0, 1]$. In words, any feasible input–output vector can be scaled down (see Figure 5.B.5). Note that nonincreasing returns to scale imply that inaction is possible [property (iv)].

(viii) *Nondecreasing returns to scale*. In contrast with the previous case, the production process exhibits nondecreasing returns to scale if for any $y \in Y$, we have $\alpha y \in Y$ for any scale $\alpha \geq 1$. In words, any feasible input–output vector can be scaled up. Figure 5.B.6(a) presents a typical example; in the figure, units of output (good 2) can be produced at a constant cost of input (good 1) except that in order to produce at all, a fixed setup cost is required. It does not matter for the existence of nondecreasing returns if this fixed cost is sunk [as in Figure 5.B.6(b)] or not [as in Figure 5.B.6(a), where inaction is possible].

(ix) *Constant returns to scale*. This property is the conjunction of properties (vii) and (viii). The production set Y exhibits constant returns to scale if $y \in Y$ implies $\alpha y \in Y$ for any scalar $\alpha \geq 0$. Geometrically, Y is a *cone* (see Figure 5.B.7).

For single-output technologies, properties of the production set translate readily into properties of the production function $f(\cdot)$. Consider Exercise 5.B.2 and Example 5.B.3.

Exercise 5.B.2: Suppose that $f(\cdot)$ is the production function associated with a single-output technology, and let Y be the production set of this technology. Show that Y satisfies constant returns to scale if and only if $f(\cdot)$ is homogeneous of degree one.



(a) (b)

Figure 5.B.6 (left)
The nondecreasing returns to scale property.

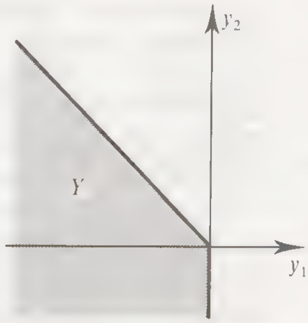


Figure 5.B.7 (right)
A technology satisfying the constant returns to scale property.

Example 5.B.3: *Returns to Scale with the Cobb–Douglas Production Function*: For the Cobb–Douglas production function introduced in Example 5.B.2, $f(2z_1, 2z_2) = 2^{\alpha+\beta}z_1^{\alpha}z_2^{\beta} = 2^{\alpha+\beta}f(z_1, z_2)$. Thus, when $\alpha + \beta = 1$, we have constant returns to scale; when $\alpha + \beta < 1$, we have decreasing returns to scale; and when $\alpha + \beta > 1$, we have increasing returns to scale. ■

(x) *Additivity (or free entry)*. Suppose that $y \in Y$ and $y' \in Y$. The additivity property requires that $y + y' \in Y$. More succinctly, $Y + Y \subset Y$. This implies, for example, that $ky \in Y$ for any positive *integer* k. In Figure 5.B.8, we see an example where Y is additive. Note that in this example, output is available only in integer amounts (perhaps because of indivisibilities). The economic interpretation of the additivity condition is that if y and y' are both possible, then one can set up two plants that do not interfere with each other and carry out production plans y and y' independently. The result is then the production vector $y + y'$.

Additivity is also related to the idea of entry. If $y \in Y$ is being produced by a firm and another firm enters and produces $y' \in Y$, then the net result is the vector $y + y'$. Hence, the *aggregate production set* (the production set describing feasible production plans for the economy as a whole) must satisfy additivity whenever unrestricted entry, or (as it is called in the literature) *free entry*, is possible.

(xi) *Convexity*. This is one of the fundamental assumptions of microeconomics. It postulates that the production set Y is convex. That is, if $y, y' \in Y$ and $\alpha \in [0, 1]$, then $\alpha y + (1 - \alpha)y' \in Y$. For example, Y is convex in Figure 5.B.5(a) but is not convex in Figure 5.B.5(b).

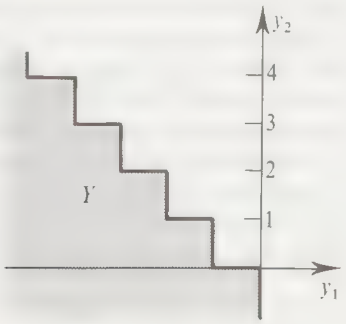


Figure 5.B.8
A production set satisfying the additivity property.

The convexity assumption can be interpreted as incorporating two ideas about production possibilities. The first is nonincreasing returns. In particular, if inaction is possible (i.e., if $0 \in Y$), then convexity implies that Y has nonincreasing returns to scale. To see this, note that for any $\alpha \in [0, 1]$, we can write $\alpha y = \alpha y + (1 - \alpha)0$. Hence, if $y \in Y$ and $0 \in Y$, convexity implies that $\alpha y \in Y$. Second, convexity captures the idea that "unbalanced" input combinations are not more productive than balanced ones (or, symmetrically, that "unbalanced" output combinations are not least costly to produce than balanced ones). In particular, if production plans y and y' produce exactly the same amount of output but use different input combinations, then a production vector that uses a level of each input that is the average of the levels used in these two plans can do at least as well as either y or y'.

Exercise 5.B.3 illustrates these two ideas for the case of a single-output technology.

Exercise 5.B.3: Show that for a single-output technology, Y is convex if and only if the production function $f(z)$ is concave.

(xii) *Y is a convex cone.* This is the conjunction of the convexity (xi) and constant returns to scale (ix) properties. Formally, Y is a convex cone if for any production vector $y, y' \in Y$ and constants $\alpha \geq 0$ and $\beta \geq 0$, we have $\alpha y + \beta y' \in Y$. The production set depicted in Figure 5.B.7 is a convex cone.

An important fact is given in Proposition 5.B.1.

Proposition 5.B.1: The production set Y is additive and satisfies the nonincreasing returns condition if and only if it is a convex cone.

Proof: The definition of a convex cone directly implies the nonincreasing returns and additivity properties. Conversely, we want to show that if nonincreasing returns and additivity hold, then for any $y, y' \in Y$ and any $\alpha > 0$, and $\beta > 0$, we have $\alpha y + \beta y' \in Y$. To this effect, let k be any integer such that $k > \text{Max}\,\{\alpha, \beta\}$. By additivity, $ky \in Y$ and $ky' \in Y$. Since $(\alpha/k) < 1$ and $\alpha y = (\alpha/k)ky$, the nonincreasing returns condition implies that $\alpha y \in Y$. Similarly, $\beta y' \in Y$. Finally, again by additivity, $\alpha y + \beta y' \in Y$. ■

Proposition 5.B.1 provides a justification for the convexity assumption in production. Informally, we could say that if feasible input–output combinations can always be scaled down, and if the simultaneous operation of several technologies without mutual interference is always possible, then, in particular, convexity obtains. (See Appendix A of Chapter 11 for several examples in which there is mutual interference and, as a consequence, convexity does not arise.)

It is important not to lose sight of the fact that the production set describes technology, not limits on resources. It can be argued that if all inputs (including, say, entrepreneurial inputs) are explicitly accounted for, then it should always be possible to replicate production. After all, we are not saying that doubling output is actually feasible, only that in principle it would be possible if *all* inputs (however esoteric, be they marketed or not) were doubled. In this view, which originated with Marshall and has been much emphasized by McKenzie (1959), decreasing returns must reflect the scarcity of an underlying, unlisted input of production. For this reason, some economists believe that among models with convex technologies the constant returns model is the most fundamental. Proposition 5.B.2 makes this idea precise.

Proposition 5.B.2: For any convex production set $Y \subset \mathbb{R}^L$ with $0 \in Y$, there is a constant returns, convex production set $Y' \subset \mathbb{R}^{L+1}$ such that $Y = \{y \in \mathbb{R}^L \colon (y, -1) \in Y'\}$.

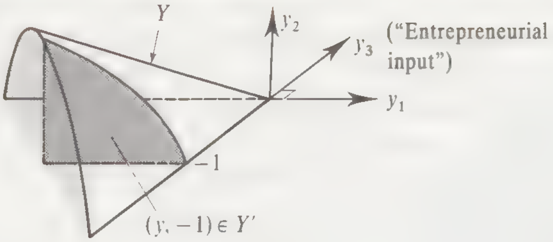


Figure 5.B.9
A constant returns production set with an "entrepreneurial factor."

Proof: Simply let $Y' = \{y' \in \mathbb{R}^{L+1}: y' = \alpha(y, -1)$ for some $y \in Y$ and $\alpha \geq 0\}$. (See Figure 5.B.9.) ■

The additional input included in the extended production set (good $L + 1$) can be called the "entrepreneurial factor." (The justification for this can be seen in Exercise 5.C.12; in a competitive environment, the return to this entrepreneurial factor is precisely the firm's profit.) In essence, the implication of Proposition 5.B.2 is that in a competitive, convex setting, there may be little loss of conceptual generality in limiting ourselves to constant returns technologies.

5.C Profit Maximization and Cost Minimization

In this section, we begin our study of the market behavior of the firm. In parallel to our study of consumer demand, we assume that there is a vector of prices quoted for the L goods, denoted by $p = (p_1, \ldots, p_L) \gg 0$, and that these prices are independent of the production plans of the firm (the *price-taking assumption*).

We assume throughout this chapter that the firm's objective is to maximize its profit. (It is quite legitimate to ask why this should be so, and we will offer a brief discussion of the issue in Section 5.G.) Moreover, we always assume that the firm's production set Y satisfies the properties of *nonemptiness*, *closedness*, and *free disposal* (see Section 5.B).

The Profit Maximization Problem

Given a price vector $p \gg 0$ and a production vector $y \in \mathbb{R}^L$, the profit generated by implementing y is $p \cdot y = \sum_{\ell=1}^{L} p_\ell y_\ell$. By the sign convention, this is precisely the total revenue minus the total cost. Given the technological constraints represented by its production set Y, the firm's *profit maximization problem* (*PMP*) is then

$$\begin{aligned} \underset{y}{\text{Max}} \quad & p \cdot y \\ \text{s.t. } & y \in Y. \end{aligned} \qquad \text{(PMP)}$$

Using a transformation function to describe Y, $F(\cdot)$, we can equivalently state the PMP as

$$\begin{aligned} \underset{y}{\text{Max}} \quad & p \cdot y \\ \text{s.t. } & F(y) \leq 0. \end{aligned}$$

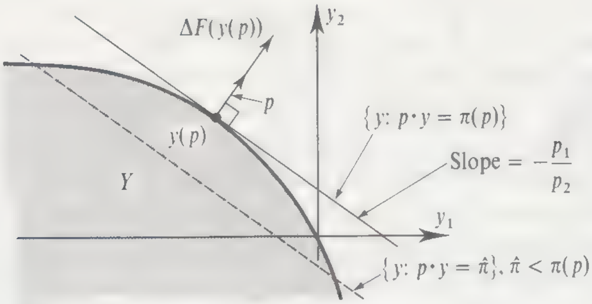


Figure 5.C.1
The profit maximization pr

Given a production set Y, the firm's *profit function* $\pi(p)$ associates to every p the amount $\pi(p) = \text{Max}\,\{p \cdot y: y \in Y\}$, the value of the solution to the PMP. Correspondingly, we define the firm's *supply correspondence* at p, denoted $y(p)$, as the set of profit-maximizing vectors $y(p) = \{y \in Y: p \cdot y = \pi(p)\}$.[4] Figure 5.C.1 depicts the supply to the PMP for a strictly convex production set Y. The optimizing vector $y(p)$ lies at the point in Y associated with the highest level of profit. In the figure, $y(p)$ therefore lies on the *iso-profit line* (a line in $\mathbb{R}^2$ along which all points generate equal profits) that intersects the production set farthest to the northeast and is, therefore, tangent to the boundary of Y at $y(p)$.

In general, $y(p)$ may be a set rather than a single vector. Also, it is possible that no profit-maximizing production plan exists. For example, the price system may be such that there is no bound on how high profits may be. In this case, we say that $\pi(p) = +\infty$.[5] To take a concrete example, suppose that $L = 2$ and that a firm with a constant returns technology produces one unit of good 2 for every unit of good 1 used as an input. Then $\pi(p) = 0$ whenever $p_2 \le p_1$. But if $p_2 > p_1$, then the firm's profit is $(p_2 - p_1)y_2$, where y_2 is the production of good 2. Clearly, by choosing y_2 appropriately, we can make profits arbitrarily large. Hence, $\pi(p) = +\infty$ if $p_2 > p_1$.

Exercise 5.C.1: Prove that, in general, if the production set Y exhibits nondecreasing returns to scale, then either $\pi(p) \le 0$ or $\pi(p) = +\infty$.

If the transformation function $F(\cdot)$ is differentiable, then first-order conditions can be used to characterize the solution to the PMP. If $y^* \in y(p)$, then, for some $\lambda \ge 0$, y^* must satisfy the first-order conditions

$$p_\ell = \lambda \frac{\partial F(y^*)}{\partial y_\ell} \quad \text{for } \ell = 1, \dots, L$$

or, equivalently, in matrix notation,

$$p = \lambda \nabla F(y^*). \tag{5.C.1}$$

4. We use the term *supply correspondence* to keep the parallel with the *demand* terminology of the consumption side. Recall however that $y(p)$ is more properly thought of as the firm's *net* supply to the market. In particular, the negative entries of a supply vector should be interpreted as demand for inputs.

5. Rigorously, to allow for the possibility that $\pi(p) = +\infty$ (as well as for other cases where no profit-maximizing production plan exists), the profit function should be defined by $\pi(p) = \text{Sup}\,\{p \cdot y: y \in Y\}$. We will be somewhat loose, however, and continue to use *Max* while allowing for this possiblity.

In words, the *price vector p and the gradient $\nabla F(y^*)$ are proportional* (Figure 5.C.1 depicts this fact). Condition (5.C.1) also yields the following ratio equality: $p_\ell / p_k = MRT_{\ell k}(y^*)$ for all ℓ, k. For $L = 2$, this says that the slope of the transformation frontier at the profit-maximizing production plan must be equal to the negative of the price ratio, as shown in Figure 5.C.1. Were this not so, a small change in the firm's production plan could be found that increases the firm's profits.

When Y corresponds to a single-output technology with differentiable production function $f(z)$, we can view the firm's decision as simply a choice over its input levels z. In this special case, we shall let the scalar $p > 0$ denote the price of the firm's output and the vector $w \gg 0$ denote its input prices.[6] The input vector z^* maximizes profit given (p, w) if it solves

$$\operatorname*{Max}_{z \geq 0} \; pf(z) - w \cdot z.$$

If z^* is optimal, then the following first-order conditions must be satisfied for $\ell = 1, \ldots, L-1$:

$$p \frac{\partial f(z^*)}{\partial z_\ell} \leq w_\ell, \quad \text{with equality if } z_\ell^* > 0,$$

or, in matrix notation,

$$p\,\nabla f(z^*) \leq w \qquad \text{and} \qquad [p\,\nabla f(z^*) - w] \cdot z^* = 0.^7 \tag{5.C.2}$$

Thus, the marginal product of every input ℓ actually used (i.e., with $z_\ell^* > 0$) must equal its price in terms of output, w_ℓ / p. Note also that for any two inputs ℓ and k with $(z_\ell^*, z_k^*) \gg 0$, condition (5.C.2) implies that $MRTS_{\ell k} = w_\ell / w_k$; that is, the marginal rate of technical substitution between the two inputs is equal to their price ratio, the economic rate of substitution between them. This ratio condition is merely a special case of the more general condition derived in (5.C.1).

If the production set Y is convex, then the first-order conditions in (5.C.1) and (5.C.2) are not only necessary but also sufficient for the determination of a solution to the PMP.

Proposition 5.C.1, which lists the properties of the profit function and supply correspondence, can be established using methods similar to those we employed in Chapter 3 when studying consumer demand. Observe, for example, that mathematically the concept of the profit function should be familiar from the discussion of duality in Chapter 3. In fact, $\pi(p) = -\mu_{-Y}(p)$, where $\mu_{-Y}(p) = \text{Min}\{p \cdot (-y) \colon y \in Y\}$ is the support function of the set $-Y$. Thus, the list of important properties in Proposition 5.C.1 can be seen to follow from the general properties of support functions discussed in Section 3.F.

6. Up to now, we have always used the symbol p for an overall vector of prices; here we use it only for the output price and we denote the vector of input prices by w. This notation is fairly standard. As a rule of thumb, unless we are in a context of explicit classification of commodities as inputs or outputs (as in the single-output case), we will continue to use p to denote an overall vector of prices $p = (p_1, \ldots, p_L)$.

7. The concern over boundary conditions arises here, but not in condition (5.C.1), because the assumption of distinct inputs and outputs requires that $z \geq 0$, whereas the formulation leading to (5.C.1) allows the net output of every good to be either positive or negative. Nonetheless, when using the first-order conditions (5.C.2), we will typically assume that $z^* \gg 0$.

Proposition 5.C.1: Suppose that $\pi(\cdot)$ is the profit function of the production set Y and that $y(\cdot)$ is the associated supply correspondence. Assume also that Y is closed and satisfies the free disposal property. Then

(i) $\pi(\cdot)$ is homogeneous of degree one.

(ii) $\pi(\cdot)$ is convex.

(iii) If Y is convex, then $Y = \{y \in \mathbb{R}^L : p \cdot y \leq \pi(p) \text{ for all } p \gg 0\}$.

(iv) $y(\cdot)$ is homogeneous of degree zero.

(v) If Y is convex, then $y(p)$ is a convex set for all p. Moreover, if Y is strictly convex, then $y(p)$ is single-valued (if nonempty).

(vi) (*Hotelling's lemma*) If $y(\bar{p})$ consists of a single point, then $\pi(\cdot)$ is differentiable at $\bar{p}$ and $\nabla\pi(\bar{p}) = y(\bar{p})$.

(vii) If $y(\cdot)$ is a function differentiable at $\bar{p}$, then $Dy(\bar{p}) = D^2\pi(\bar{p})$ is a symmetric and positive semidefinite matrix with $Dy(\bar{p})\bar{p} = 0$.

Properties (ii), (iii), (vi), and (vii) are the nontrivial ones.

Exercise 5.C.2: Prove that $\pi(\cdot)$ is a convex function [Property (ii) of Proposition 5.C.1]. [*Hint*: Suppose that $y \in y(\alpha p + (1-\alpha)p')$. Then

$$\pi(\alpha p + (1-\alpha)p') = \alpha p \cdot y + (1-\alpha)p' \cdot y \leq \alpha\pi(p) + (1-\alpha)\pi(p').]$$

Property (iii) tells us that if Y is closed, convex, and satisfies free disposal, then $\pi(p)$ provides an alternative ("dual") description of the technology. As for the indirect utility function's (or expenditure function's) representation of preferences (discussed in Chapter 3), it is a less primitive description than Y itself because it depends on the notions of prices and of price-taking behavior. But thanks to property (vi), it has the great virtue in applications of often allowing for an immediate computation of supply.

Property (vi) relates supply behavior to the derivatives of the profit function. It is a direct consequence of the duality theorem (Proposition 3.F.1). As in Proposition 3.G.1, the fact that $\nabla\pi(\bar{p}) = y(\bar{p})$ can also be established by the related arguments of the envelope theorem and of first-order conditions.

The positive semidefiniteness of the matrix $Dy(p)$ in property (vii), which in view of property (vi) is a consequence of the convexity of $\pi(\cdot)$, is the general mathematical expression of the *law of supply*: *Quantities respond in the same direction as price changes*. By the sign convention, this means that *if the price of an output increases* (all other prices remaining the same), then *the supply of the output increases*; and *if the price of an input increases*, then *the demand for the input decreases.*

Note that the law of supply holds for *any* price change. Because, in contrast with demand theory, there is no budget constraint, there is no compensation requirement of any sort. In essence, we have no wealth effects here, only substitution effects.

In nondifferentiable terms, the law of supply can be expressed as

$$(p - p') \cdot (y - y') \geq 0 \tag{5.C.3}$$

for all p, p', $y \in y(p)$, and $y' \in y(p')$. In this form, it can also be established by a straightforward revealed preference argument. In particular,

$$(p - p') \cdot (y - y') = (p \cdot y - p \cdot y') + (p' \cdot y' - p' \cdot y) \geq 0,$$

where the inequality follows from the fact that $y \in y(p)$ and $y' \in y(p')$ (i.e., from the fact that y is profit maximizing given prices p and y' is profit maximizing for prices p').

Property (vii) of Proposition 5.C.1 implies that the matrix $Dy(p)$, the *supply substitution matrix*, has properties that parallel (although with the reverse sign) those for the substitution matrix of demand theory. Thus, own-substitution effects are nonnegative as noted above $[\partial y_\ell(p)/\partial p_\ell \geq 0$ for all $\ell]$, and substitution effects are symmetric $[\partial y_\ell(p)/\partial p_k = \partial y_k(p)/\partial p_\ell$ for all $\ell, k]$. The fact that $Dy(p)p = 0$ follows from the homogeneity of $y(\cdot)$ [property (iv)] in a manner similar to the parallel property of the demand substitution matrix discussed in Chapter 3.

Cost Minimization

An important implication of the firm choosing a profit-maximizing production plan is that there is no way to produce the same amounts of outputs at a lower total input cost. Thus, cost minimization is a necessary condition for profit maximization. This observation motivates us to an independent study of the firm's *cost minimization problem*. The problem is of interest for several reasons. First, it leads us to a number of results and constructions that are technically very useful. Second, as we shall see in Chapter 12, when a firm is not a price taker in its output market, we can no longer use the profit function for analysis. Nevertheless, as long as the firm is a price taker in its input market, the results flowing from the cost minimization problem continue to be valid. Third, when the production set exhibits nondecreasing returns to scale, the value function and optimizing vectors of the cost minimization problem, which keep the levels of outputs fixed, are better behaved than the profit function and supply correspondence of the PMP (e.g., recall from Exercise 5.C.1 that the profit function can take only the values 0 and $+\infty$).

To be concrete, we focus our analysis on the single-output case. As usual, we let z be a nonnegative vector of inputs, $f(z)$ the production function, q the amounts of output, and $w \gg 0$ the vector of input prices. The *cost minimization problem* (CMP) can then be stated as follows (we assume free disposal of output):

$$\begin{aligned} \underset{z \geq 0}{\text{Min}} \quad & w \cdot z \\ \text{s.t.} \quad & f(z) \geq q. \end{aligned} \tag{CMP}$$

The optimized value of the CMP is given by the *cost function* $c(w, q)$. The corresponding optimizing set of input (or factor) choices, denoted by $z(w, q)$, is known as the *conditional factor demand correspondence* (or *function* if it is always single-valued). The term *conditional* arises because these factor demands are conditional on the requirement that the output level q be produced.

The solution to the CMP is depicted in Figure 5.C.2(a) for a case with two inputs. The shaded region represents the set of input vectors z that can produce at least the amount q of output. It is the projection (into the positive orthant of the input space) of the part of the production set Y than generates output of at least q, as shown in Figure 5.C.2(b). In Figure 5.C.2(a), the solution $z(w, q)$ lies on the iso-cost line (a line in $\mathbb{R}^2$ on which all input combinations generate equal cost) that intersects the set $\{z \in \mathbb{R}^L_+ : f(z) \geq q\}$ closest to the origin.

If z^* is optimal in the CMP, and if the production function $f(\cdot)$ is differentiable,

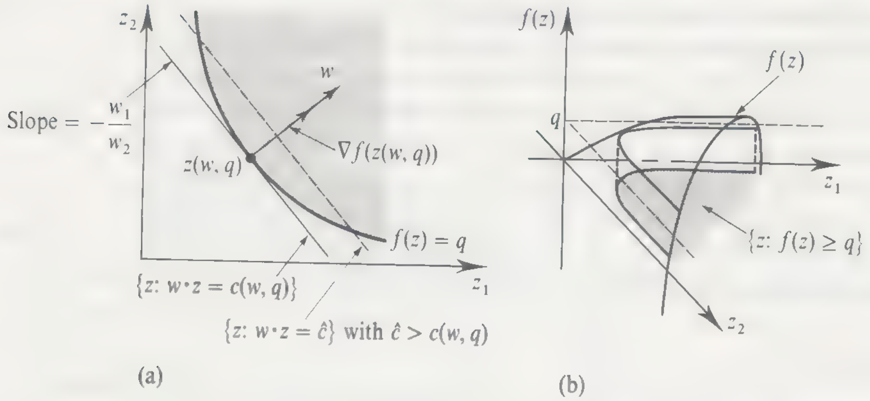


Figure 5.C.2
The cost minimization problem.
(a) Two inputs.
(b) The isoquant as a section of the production set.

then for some $\lambda \geq 0$, the following first-order conditions must hold for every input $\ell = 1, \ldots, L-1$:

$$w_\ell \geq \lambda \frac{\partial f(z^*)}{\partial z_\ell}, \quad \text{with equality if } z_\ell^* > 0,$$

or, in matrix notation,

$$w \geq \lambda \nabla f(z^*) \quad \text{and} \quad [w - \lambda \nabla f(z^*)] \cdot z^* = 0. \tag{5.C.4}$$

As with the PMP, if the production set Y is convex [i.e., if $f(\cdot)$ is concave], then condition (5.C.4) is not only necessary but also sufficient for z^* to be an optimum in the CMP.[8]

Condition (5.C.4), like condition (5.C.2) of the PMP, implies that for any two inputs ℓ and k with $(z_\ell, z_k) \gg 0$, we have $MRTS_{\ell k} = w_\ell/w_k$. This correspondence is to be expected because, as we have noted, profit maximization implies that input choices are cost minimizing for the chosen output level q. For $L = 2$, condition (5.C.4) entails that the slope at z^* of the *isoquant* associated with production level q is exactly equal to the negative of the ratio of the input prices $-w_1/w_2$. Figure 5.C.2(a) depicts this fact as well.

As usual, the Lagrange multiplier λ can be interpreted as the marginal value of relaxing the constraint $f(z^*) \geq q$. Thus, λ equals $\partial c(w, q)/\partial q$, the *marginal cost of production.*

Note the close formal analogy with consumption theory here. Replace $f(\cdot)$ by $u(\cdot)$, q by u, and z by x (i.e., interpret the production function as a utility function), and the CMP becomes the expenditure minimization problem (EMP) discussed in Section 3.E. Therefore, in Proposition 5.C.2, properties (i) to (vii) of the cost function and conditional factor demand correspondence follow from the analysis in Sections 3.E to 3.G by this reinterpretation. [You are asked to prove properties (viii) and (ix) in Exercise 5.C.3.]

Proposition 5.C.2: Suppose that $c(w, q)$ is the cost function of a single-output technology Y with production function $f(\cdot)$ and that $z(w, q)$ is the associated

8. Note, however, that the first-order conditions are sufficient for a solution to the CMP as long as the set $\{z: f(z) \geq q\}$ is convex. Thus, the key condition for the sufficiency of the first-order conditions of the CMP is the *quasiconcavity* of $f(\cdot)$. This is an important fact because the quasiconcavity of $f(\cdot)$ is compatible with increasing returns to scale (see Example 5.C.1).

conditional factor demand correspondence. Assume also that Y is closed and satisfies the free disposal property. Then

(i) $c(\cdot)$ is homogeneous of degree one in w and nondecreasing in q.
(ii) $c(\cdot)$ is a concave function of w.
(iii) If the sets $\{z \geq 0: f(z) \geq q\}$ are convex for every q, then $Y = \{(-z, q): w \cdot z \geq c(w, q) \text{ for all } w \gg 0\}$.
(iv) $z(\cdot)$ is homogeneous of degree zero in w.
(v) If the set $\{z \geq 0: f(z) \geq q\}$ is convex, then $z(w, q)$ is a convex set. Moreover, if $\{z \geq 0: f(z) \geq q\}$ is a strictly convex set, then $z(w, q)$ is single-valued.
(vi) (*Shepard's lemma*) If $z(\bar{w}, q)$ consists of a single point, then $c(\cdot)$ is differentiable with respect to w at $\bar{w}$ and $\nabla_w c(\bar{w}, q) = z(\bar{w}, q)$.
(vii) If $z(\cdot)$ is differentiable at $\bar{w}$, then $D_w z(\bar{w}, q) = D_w^2 c(\bar{w}, q)$ is a symmetric and negative semidefinite matrix with $D_w z(\bar{w}, q)\bar{w} = 0$.
(viii) If $f(\cdot)$ is homogeneous of degree one (i.e., exhibits constant returns to scale), then $c(\cdot)$ and $z(\cdot)$ are homogeneous of degree one in q.
(ix) If $f(\cdot)$ is concave, then $c(\cdot)$ is a convex function of q (in particular, marginal costs are nondecreasing in q).

In Exercise 5.C.4 we are asked to show that properties (i) to (vii) of Proposition 5.C.2 also hold for technologies with multiple outputs.

The cost function can be particularly useful when the production set is of the constant returns type. In this case, $y(\cdot)$ is not single-valued at any price vector allowing for nonzero production, making Hotelling's lemma [Proposition 5.C.1(vi)] inapplicable at these prices. Yet, the conditional input demand $z(w, q)$ may nevertheless be single-valued, allowing us to use Shepard's lemma. Keep in mind, however, that the cost function does not contain more information than the profit function. In fact, we know from property (iii) of Propositions 5.C.1 and 5.C.2 that under convexity restrictions there is a one-to-one correspondence between profit and cost functions; that is, from either function, the production set can be recovered, and the other function can then be derived.

Using the cost function, we can restate the firm's problem of determining its profit-maximizing production level as

$$\underset{q \geq 0}{\text{Max}}\ pq - c(w, q). \tag{5.C.5}$$

The necessary first-order condition for q^* to be profit maximizing is then

$$p - \frac{\partial c(w, q^*)}{\partial q} \leq 0, \quad \text{with equality if } q^* > 0. \tag{5.C.6}$$

In words, at an interior optimum (i.e., if $q^* > 0$), *price equals marginal cost*.[9] If $c(w, q)$ is convex in q, then the first-order condition (5.C.6) is also sufficient for q^* to be the firm's optimal output level. (We study the relationship between the firm's supply behavior and the properties of its technology and cost function in detail in Section 5.D.)

9. This can also be seen by noting that the first-order condition (5.C.4) of the CMP coincides with first-order condition (5.C.2) of the PMP if and only if $\lambda = p$. Recall that λ, the multiplier on the constraint in the CMP, is equal to $\partial c(w, q)/\partial q$.

We could go on for many pages analyzing profit and cost functions. Some examples and further properties are contained in the exercises. See McFadden (1978) for an extensive treatment of this topic.

Example 5.C.1: *Profit and Cost Functions for the Cobb Douglas Production Function.* Here we derive the profit and cost functions for the Cobb–Douglas production function of Example 5.B.2, $f(z_1, z_2) = z_1^\alpha z_2^\beta$. Recall from Example 5.B.3 that $\alpha + \beta = 1$ corresponds to the case of constant returns to scale, $\alpha + \beta < 1$ corresponds to decreasing returns, and $\alpha + \beta > 1$ corresponds to increasing returns.

The conditional factor demand equations and cost function have exactly the same form, and are derived in exactly the same way, as the expenditure function in Section 3.E (see Example 3.E.1; the only difference in the computations is that we now do not impose $\alpha + \beta = 1$):

$$z_1(w_1, w_2, q) = q^{1/(\alpha+\beta)}(\alpha w_2/\beta w_1)^{\beta/(\alpha+\beta)},$$
$$z_2(w_1, w_2, q) = q^{1/(\alpha+\beta)}(\beta w_1/\alpha w_2)^{\alpha/(\alpha+\beta)},$$

and

$$c(w_1, w_2, q) = q^{1/(\alpha+\beta)}[(a/\beta)^{\beta/(\alpha+\beta)} + (\alpha/\beta)^{-\alpha/(\alpha+\beta)}]w_1^{\alpha/(\alpha+\beta)}w_2^{\beta/(\alpha+\beta)}.$$

This cost function has the form $c(w_1, w_2, q) = q^{1/(\alpha+\beta)}\theta\phi(w_1, w_2)$, where

$$\theta = [(\alpha/\beta)^{\beta/(\alpha+\beta)} + (\alpha/\beta)^{-\alpha/(\alpha+\beta)}]$$

is a constant and $\phi(w_1, w_2) = w_1^{\alpha/(\alpha+\beta)}w_2^{\beta/(\alpha+\beta)}$ is a function that does not depend on the output level q. When we have constant returns, $\theta\phi(w_1, w_2)$ is the per-unit cost of production.

One way to derive the firm's supply function and profit function is to use this cost function and solve problem (5.C.5). Applying (5.C.6), the first-order condition for this problem is

$$p \leq \theta\phi(w_1, w_2)\left(\frac{1}{\alpha+\beta}\right)q^{(1/(\alpha+\beta))-1}, \quad \text{with equality if } q > 0 \qquad (5.C.7)$$

The first-order condition (5.C.7) is sufficient for a maximum when $\alpha + \beta \leq 1$ because the firm's cost function is then convex in q.

When $\alpha + \beta < 1$, (5.C.7) can be solved for a unique optimal output level:

$$q(w_1, w_2, p) = (\alpha + \beta)[p/\theta\phi(w_1, w_2)]^{(\alpha+\beta)/(1-\alpha-\beta)}.$$

The factor demands can then be obtained through substitution,

$$z_\ell(w_1, w_2, p) = z_\ell(w_1, w_2, q(w_1, w_2, p)) \quad \text{for } \ell = 1, 2,$$

as can the profit function,

$$\pi(w_1, w_2, p) = pq(w_1, w_2, p) - w \cdot z(w_1, w_2, q(w_1, w_2, p)).$$

When $\alpha + \beta = 1$, the right-hand side of the first-order condition (5.C.7) becomes $\theta\phi(w_1, w_2)$, the unit cost of production (which is independent of q). If $\theta\phi(w_1, w_2)$ is greater than p, then $q = 0$ is optimal; if it is smaller than p, then no solution exists (again, unbounded profits can be obtained by increasing q); and when $\theta\phi(w_1, w_2) = p$, any non-negative output level is a solution to the PMP and generates zero profits.

Finally, when $\alpha + \beta > 1$ (so that we have increasing returns to scale), a quantity q satisfying the first-order condition (5.C.7) does not yield a profit-maximizing production. [Actually, in this case, the cost function is strictly concave in q, so that

any solution to the first-order condition (5.C.7) yields a local *minimum* of profits, subject to output being always produced at minimum cost]. Indeed, since $p > 0$, a doubling of the output level starting from any q doubles the firm's revenue but increases input costs only by a factor of $2^{1/(\alpha+\beta)} > 2$. With enough doublings, the firm's profits can therefore be made arbitrarily large. Hence, with increasing returns to scale, there is no solution to the PMP. ■

5.D The Geometry of Cost and Supply in the Single-Output Case

In this section, we continue our analysis of the relationships among a firm's technology, its cost function, and its supply behavior for the special but commonly used case in which there is a single output. A significant advantage of considering the single-output case is that it lends itself to extensive graphical illustration.

Throughout, we denote the amount of output by q and hold the vector of factor prices constant at $\bar{w} \gg 0$. For notational convenience, we write the firm's cost function as $C(q) = c(\bar{w}, q)$. For $q > 0$, we can denote the firm's average cost by $AC(q) = C(q)/q$ and assuming that the derivative exists, we denote its *marginal cost* by $C'(q) = dC(q)/dq$.

Recall from expression (5.C.6) that for a given output price p, all profit-maximizing output levels $q \in q(p)$ must satisfy the first-order condition [assuming that $C'(q)$ exists]:

$$p \leq C'(q) \quad \text{with equality if } q > 0. \tag{5.D.1}$$

If the production set Y is convex, $C(\cdot)$ is a convex function [see property (ix) of Proposition 5.C.2], and therefore marginal cost is nondecreasing. In this case, as we noted in Section 5.C, satisfaction of this first-order condition is also sufficient to establish that q is a profit-maximizing output level at price p.

Two examples of convex production sets are given in Figures 5.D.1 and 5.D.2. In the figures, we assume that there is only one input, and we normalize its price to equal 1 (you can think of this input as the total expense of factor use).[10] Figure 5.D.1 depicts the production set (a), cost function (b), and average and marginal cost functions (c) for a case with decreasing returns to scale. Observe that the cost function is obtained from the production set by a 90-degree rotation. The determination of average cost and marginal cost from the cost function is shown in Figure 5.D.1(b) (for an output level $\hat{q}$). Figure 5.D.2 depicts the same objects for a case with constant returns to scale.

In Figures 5.D.1(c) and 5.D.2(c), we use a heavier trace to indicate the firm's profit-maximizing supply locus, the graph of $q(\cdot)$. (*Note: In this and subsequent figures, the supply locus is always indicated by a heavier trace.*) Because the technologies in these two examples are convex, the supply locus in each case coincides exactly with the (q, p) combinations that satisfy the first-order condition (5.D.1).

If the technology is not convex, perhaps because of the presence of some underlying indivisibility, then satisfaction of the first-order necessary condition

10. Thus, the single input can be thought of as a Hicksian composite commodity in a sense analogous to that in Exercise 3.G.5.

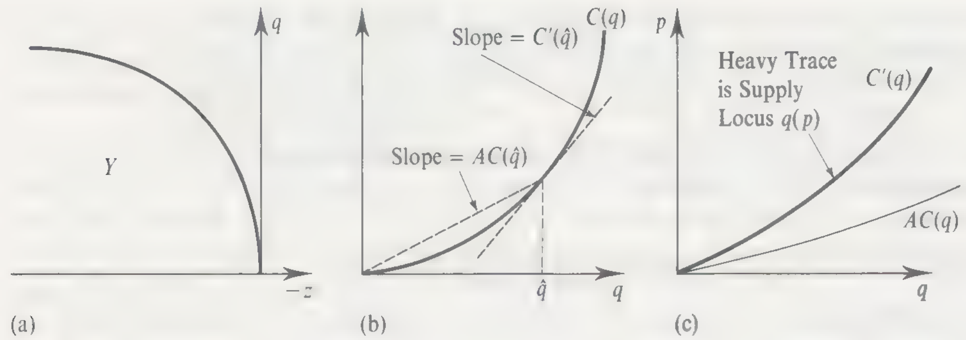


Figure 5.D.1
A strictly convex technology (strictly decreasing returns to scale).
(a) Production set.
(b) Cost function.
(c) Average cost, marginal cost, and supply.

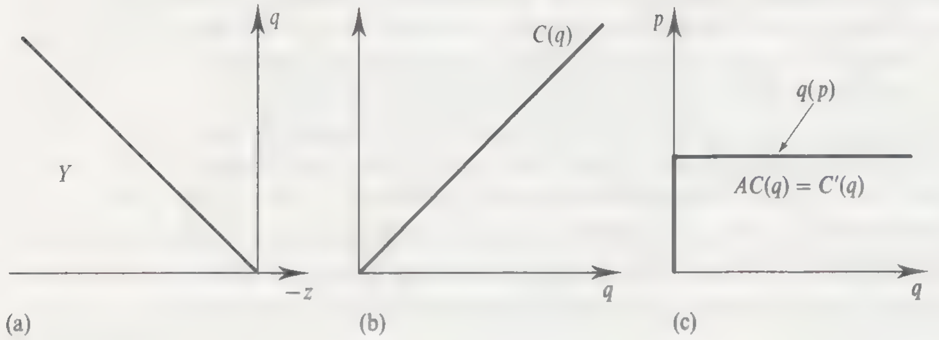


Figure 5.D.2
A constant returns to scale technology.
(a) Production set.
(b) Cost function.
(c) Average cost, marginal cost, and supply.

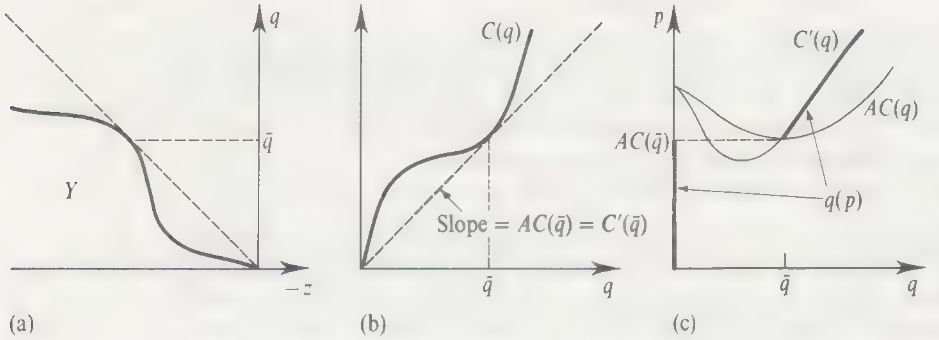


Figure 5.D.3
A nonconvex technology.
(a) Production set.
(b) Cost function.
(c) Average cost, marginal cost, and supply.

(5.D.1) no longer implies that q is profit maximizing. The supply locus will then be only a subset of the set of (q, p) combinations that satisfy (5.D.1).

Figure 5.D.3 depicts a situation with a nonconvex technology. In the figure, we have an initial segment of increasing returns over which the average cost decreases and then a region of decreasing returns over which the average cost increases. The level (or levels) of production corresponding to the minimum average cost is called the *efficient scale*, which, if unique, we denote by $\bar{q}$. Looking at the cost functions in Figure 5.D.3(a) and (b), we see that at $\bar{q}$ we have $AC(\bar{q}) = C'(\bar{q})$. In Exercise 5.D.1, you are asked to establish this fact as a general result.

Exercise 5.D.1: Show that $AC(\bar{q}) = C'(\bar{q})$ at any $\bar{q}$ satisfying $AC(\bar{q}) \leq AC(q)$ for all q. Does this result depend on the differentiability of $C(\cdot)$ everywhere?

The supply locus for this nonconvex example is depicted by the heavy trace in

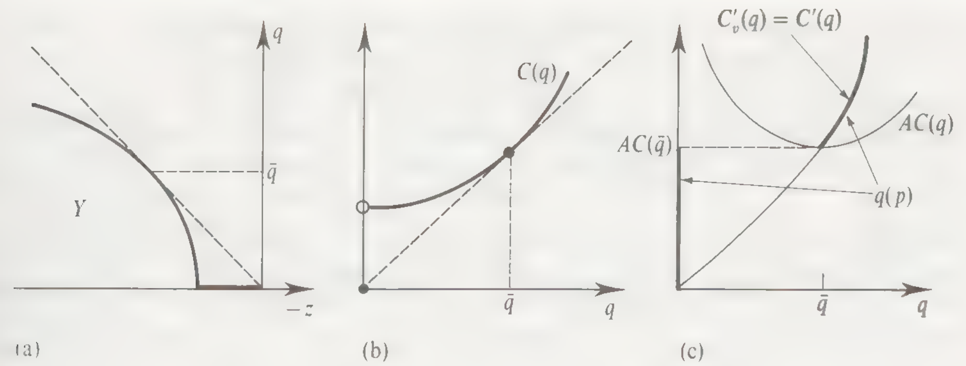


Figure 5.D.4
Strictly convex variable costs with a nonsunk setup cost. (a) Production set. (b) Cost function. (c) Average cost, marginal cost, and supply.

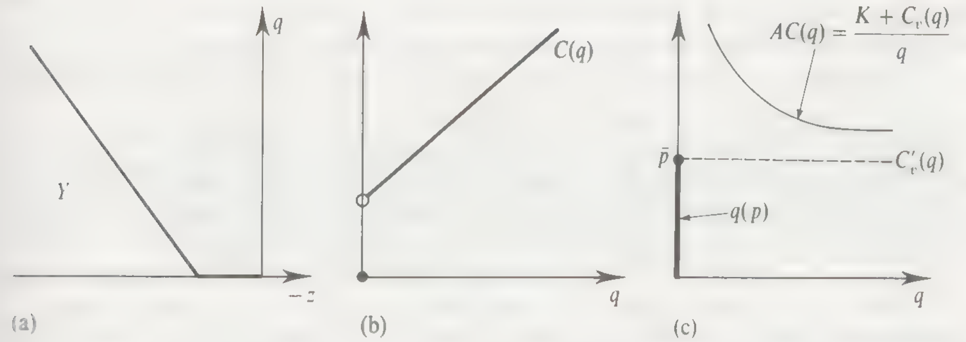


Figure 5.D.5
Constant returns variable costs with a nonsunk setup cost. (a) Production set. (b) Cost function. (c) Average cost, marginal cost, and supply.

Figure 5.D.3(c). When $p > AC(\bar{q})$, the firm maximizes its profit by producing at the unique level of q satisfying $p = C'(q) > AC(q)$. [Note that the firm earns strictly positive profits doing so, exceeding the zero profits earned by setting $q = 0$, which in turn exceed the strictly negative profits earned by choosing any $q > 0$ with $p = C'(q) < AC(q)$.] On the other hand, when $p < AC(\bar{q})$, any $q > 0$ earns strictly negative profits, and so the firm's optimal supply is $q = 0$ [note that $q = 0$ satisfies the necessary first-order condition (5.D.1) because $p < C'(0)$]. When $p = AC(\bar{q})$, the profit-maximizing set of output levels is $\{0, \bar{q}\}$. The supply locus is therefore as shown in Figure 5.D.3(c).

An important source of nonconvexities is fixed setup costs. These may or may not be sunk. Figures 5.D.4 and 5.D.5 (which parallel 5.D.1 and 5.D.2) depict two cases with nonsunk fixed setup costs (so inaction is possible). In these figures, we consider a case in which the firm incurs a fixed cost K if and only if it produces a positive amount of output and otherwise has convex costs. In particular, total cost is of the form $C(0) = 0$, and $C(q) = C_v(q) + K$ for $q > 0$, where $K > 0$ and $C_v(q)$, the *variable cost function*, is convex [and has $C_v(0) = 0$]. Figure 5.D.4 depicts the case in which $C_v(\cdot)$ is strictly convex, whereas $C_v(\cdot)$ is linear in Figure 5.D.5. The supply loci are indicated in the figures. In both illustrations, the firm will produce a positive amount of output only if its profit is sufficient to cover not only its variable costs but also the fixed cost K. You should read the supply locus in Figure 5.D.5(c) as saying that for $p > \bar{p}$, the supply is "infinite," and that $q = 0$ is optimal for $p \leq \bar{p}$.

In Figure 5.D.6, we alter the case studied in Figure 5.D.4 by making the fixed costs sunk, so that $C(0) > 0$. In particular, we now have $C(q) = C_v(q) + K$ for all $q \geq 0$; therefore, the firm must pay K whether or not it produces a positive quantity.

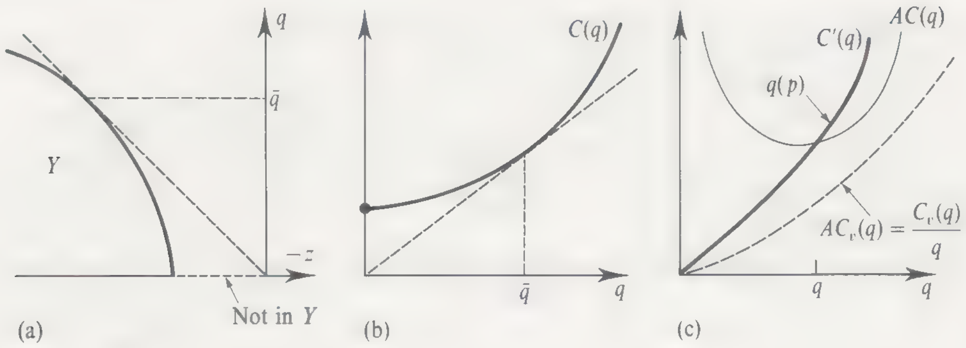


Figure 5.D.6 Strictly convex variable costs with sunk costs. (a) Production set. (b) Cost function. (c) Average cost, marginal cost, and supply.

Although inaction is not possible here, the firm's cost function is convex, and so we are back to the case in which the first-order condition (5.D.1) is sufficient. Because the firm must pay K regardless of whether it produces a positive output level, it will not shut down simply because profits are negative. Note that because $C_v(\cdot)$ is convex and $C_v(0) = 0$, $p = C_v'(q)$ implies that $pq > C_v(q)$; hence, the firm covers its variable costs when it sets output to satisfy its first-order condition. The firm's supply locus is therefore that depicted in Figure 5.D.6(c). Note that its supply behavior is exactly the same as if it did not have to pay the sunk cost K at all [compare with Figure 5.D.1(c)].

Exercise 5.D.2: Depict the supply locus for a case with partially sunk costs, that is, where $C(q) = K + C_v(q)$ if $q > 0$ and $0 < C(0) < K$.

As we noted in Section 5.B, one source of sunk costs, at least in the short run, is input choices irrevocably set by prior decisions. Suppose, for example, that we have two inputs and a production function $f(z_1, z_2)$. Recall that we keep the prices of the two inputs fixed at $(\bar{w}_1, \bar{w}_2)$. In Figure 5.D.7(a), the cost function excluding any prior input commitments is depicted by $C(\cdot)$. We call it the *long-run cost function.* If one input, say z_2, is fixed at level $\bar{z}_2$ in the short-run, then the *short-run cost function* of the firm becomes $C(q|\bar{z}_2) = \bar{w}_1 z_1 + \bar{w}_2 \bar{z}_2$, where z_1 is chosen so that $f(z_1, \bar{z}_2) = q$. Several such short-run cost functions corresponding to different levels of z_2 are illustrated in Figure 5.D.7(a). Because restrictions on the firm's input decisions can only increase its costs of production, $C(q|\bar{z}_2)$ lies above $C(q)$ at all q except the q for

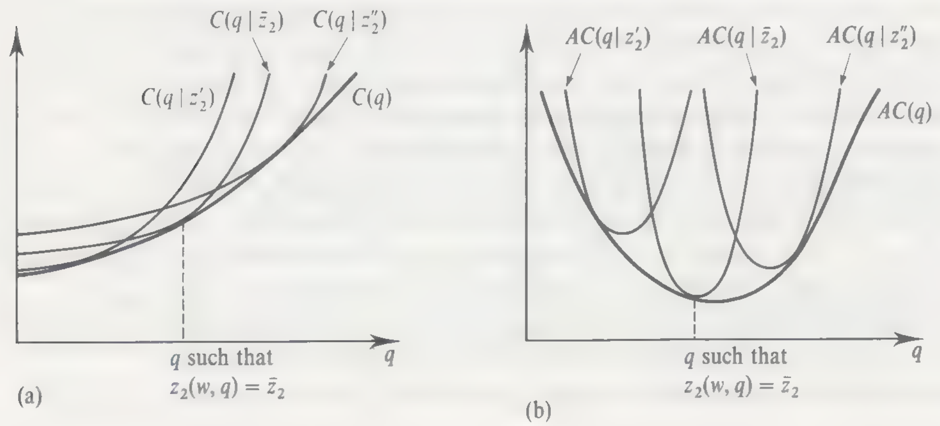


Figure 5.D.7 Costs when an input level is fixed in the short run but is free to vary in the long run. (a) Long-run and short-run cost functions. (b) Long-run and short-run average cost.

which $\bar{z}_2$ is the optimal long-run input level [i.e., the q such that $z_2(\bar{w}, q) = \bar{z}_2$]. Thus, $C(q|z_2(\bar{w},q)) = C(q)$ for all q. It follows from this and from the fact that $C(q'|z_2(\bar{w}, q)) \geq C(q')$ for all q', that $C'(q) = C'(q|z_2(\bar{w}, q))$ for all q; that is, if the level of z_2 is at its long-run value, then the short-run marginal cost equals the long-run marginal cost. Geometrically, $C(\cdot)$ is the lower envelope of the family of short-run functions $C(q|z_2)$ generated by letting z_2 take all possible values.

Observe finally that given the long-run and short-run cost functions, the long-run and short-run average cost functions and long-run and short-run supply functions of the firm can be derived in the manner discussed earlier in the section. The average-cost version of Figure 5.D.7(a) is given in Figure 5.D.7(b). (Exercise 5.D.3 asks you to investigate the short-run and long-run supply behavior of the firm in more detail.)

5.E Aggregation

In this section, we study the theory of aggregate (net) supply. As we saw in Section 5.C, the absence of a budget constraint implies that individual supply is not subject to wealth effects. As prices change, there are only substitution effects along the production frontier. In contrast with the theory of aggregate demand, this fact makes for an aggregation theory that is simple and powerful.[11]

Suppose there are J production units (firms or, perhaps, plants) in the economy, each specified by a production set $Y_1, \ldots, Y_J$. We assume that each Y_j is nonempty, closed, and satisfies the free disposal property. Denote the profit function and supply correspondences of Y_j by $\pi_j(p)$ and $y_j(p)$, respectively. The *aggregate supply correspondence* is the sum of the individual supply correspondences:

$$y(p) = \sum_{j=1}^{J} y_j(p) = \{y \in \mathbb{R}^L\colon y = \textstyle\sum_j y_j \text{ for some } y_j \in y_j(p), j = 1, \ldots, J\}.$$

Assume, for a moment, that every $y_j(\cdot)$ is a single-valued, differentiable function at a price vector p. From Proposition 5.C.1, we know that every $Dy_j(p)$ is a symmetric, positive semidefinite matrix. Because these two properties are preserved under addition, we can conclude that the matrix $Dy(p)$ is *symmetric and positive semidefinite.*

As in the theory of individual production, the positive semidefiniteness of $Dy(p)$ implies the *law of supply* in the aggregate: If a price increases, then so does the corresponding *aggregate* supply. As with the law of supply at the firm level, this property of aggregate supply holds for *all* price changes. We can also prove this aggregate law of supply directly because we know from (5.C.3) that $(p - p')\cdot[y_j(p) - y_j(p')] \geq 0$ for every j; therefore, adding over j, we get

$$(p - p')\cdot[y(p) - y(p')] \geq 0.$$

The symmetry of $Dy(p)$ suggests that underlying $y(p)$ there is a "representative producer." As we now show, this is true in a particularly strong manner.

Given $Y_1, \ldots, Y_J$, we can define the *aggregate production set* by

$$Y = Y_1 + \cdots + Y_J = \{y \in \mathbb{R}^L\colon y = \textstyle\sum_j y_j \text{ for some } y_j \in Y_j, j = 1, \ldots, J\}.$$

11. A classical and very readable account for the material in this section and in Section 5.F is Koopmans (1957).

The aggregate production set Y describes the production vectors that are feasible in the aggregate if all the production sets are used together. Let $\pi^*(p)$ and $y^*(p)$ be the profit function and the supply correspondence of the aggregate production set Y. They are the profit function and supply correspondence that would arise if a single price-taking firm were to operate, under the same management so to speak, all the individual production sets.

Proposition 5.E.1 establishes a strong aggregation result for the supply side: *The aggregate profit obtained by each production unit maximizing profit separately taking prices as given is the same as that which would be obtained if they were to coordinate their actions* (i.e., *their* y_js) *in a joint profit maximizing decision.*

Proposition 5.E.1: For all $p \gg 0$, we have

(i) $\pi^*(p) = \sum_j \pi_j(p)$

(ii) $y^*(p) = \sum_j y_j(p)$ $(=\{\sum_j y_j \colon y_j \in y_j(p)$ for every $j\})$.

Proof: (i) For the first equality, note that if we take any collection of production plans $y_j \in Y_j$, $j = 1, \dots, J$, then $\sum_j y_j \in Y$. Because $\pi^*(\cdot)$ is the profit function associated with Y, we therefore have $\pi^*(p) \geq p \cdot (\sum_j y_j) = \sum_j p \cdot y_j$. Hence, it follows that $\pi^*(p) \geq \sum_j \pi_j(p)$. In the other direction, consider any $y \in Y$. By the definition of the set Y, there are $y_j \in Y_j, j=1, \dots, J$, such that $\sum_j y_j = y$. So $p \cdot y = p \cdot (\sum_j y_j) = \sum_j p \cdot y_j \leq \sum_j \pi_j(p)$ for all $y \in Y$. Thus, $\pi^*(p) \leq \sum_j \pi_j(p)$. Together, these two inequalities imply that $\pi^*(p) = \sum_j \pi(p)$.

(ii) For the second equality, we must show that $\sum_j y_j(p) \subset y^*(p)$ and that $y^*(p) \subset \sum_j y_j(p)$. For the former relation, consider any set of individual production plans $y_j \in y_j(p), j = 1, \dots, J$. Then $p \cdot (\sum_j y_j) = \sum_j p \cdot y_j = \sum_j \pi_j(p) = \pi^*(p)$, where the last equality follows from part (i) of the proposition. Hence, $\sum_j y_j \in y^*(p)$, and therefore, $\sum_j y_j(p) \subset y^*(p)$. In the other direction, take any $y \in y^*(p)$. Then $y = \sum_j y_j$ for some $y_j \in Y_j, j = 1, \dots, J$. Since $p \cdot (\sum_j y_j) = \pi^*(p) = \sum_j \pi_j(p)$ and, for every j, we have $p \cdot y_j \leq \pi_j(p)$, it must be that $p \cdot y_j = \pi_j(p)$ for every j. Thus, $y_j \in y_j(p)$ for all j, and so $y \in \sum_j y_j(p)$. Thus, we have shown that $y^*(p) \subset \sum_j y_j(p)$. ■

The content of Proposition 5.E.1 is illustrated in Figure 5.E.1. The proposition can be interpreted as a decentralization result: To find the solution of the aggregate profit maximization problem for given prices p, it is enough to add the solutions of the corresponding individual problems.

Simple as this result may seem, it nevertheless has many important implications. Consider, for example, the single-output case. The result tells us that if firms are maximizing profit facing output price p and factor prices w, then their supply behavior maximizes aggregate profits. But this must mean that if $q = \sum_j q_j$ is the aggregate output produced by the firms, then the total cost of production is exactly equal to $c(w, q)$, the value of the *aggregate cost function* (the cost function corresponding to the aggregate production set Y). *Thus, the allocation of the production of output level q among the firms is cost minimizing.* In addition, this allows us to relate the firms' aggregate supply function for output $q(p)$ to the aggregate cost function in the same manner as done in Section 5.D for an individual firm. (This fact will prove useful when we study partial equilibrium models of competitive markets in Chapter 10.)

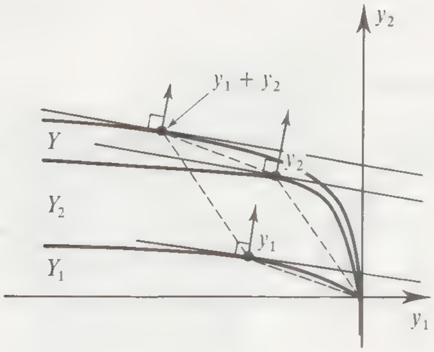


Figure 5.E.1
Joint profit maximization as a result of individual profit maximization.

In summary: If firms maximize profits taking prices as given, then the production side of the economy aggregates beautifully.

As in the consumption case (see Appendix A of Chapter 4), aggregation can also have helpful regularizing effects in the production context. An interesting and important fact is that the existence of many firms or plants with technologies that are not too dissimilar can make the *average* production set almost convex, even if the individual production sets are not so. This is illustrated in Figure 5.E.2, where there are J firms with identical production sets equal to

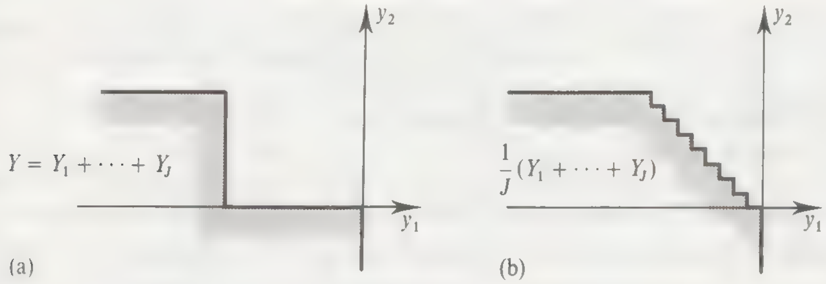


Figure 5.E.2
An example of the convexifying effects of aggregation.
(a) The individual production set.
(b) The average production set.

that displayed in 5.E.2(a). Defining the average production set as $(1/J)(Y_1 + \cdots + Y_J) = \{y: y = (1/J)(y_1 + \cdots + y_J) \text{ for some } y_j \in Y_j, j = 1, \ldots, J\}$, we see that for large J, this set is nearly convex, as depicted in Figure 5.E.2(b).[12]

5.F Efficient Production

Because much of welfare economics focuses on efficiency (see, for example, Chapters 10 and 16), it is useful to have algebraic and geometric characterizations of productions plans that can unambiguously be regarded as nonwasteful. This motivates Definition 5.F.1.

12. Note that this production set is bounded above. This is important because it insures that the individual nonconvexity is of finite size. If the individual production set was like that shown in, say, Figure 5.B.4, where neither the set nor the nonconvexity is bounded, then the average set would display a large nonconvexity (for any J). In Figure 5.B.5, we have a case of an unbounded production set but with a bounded nonconvexity; as for Figure 5.E.2, the average set will in this case be almost convex.

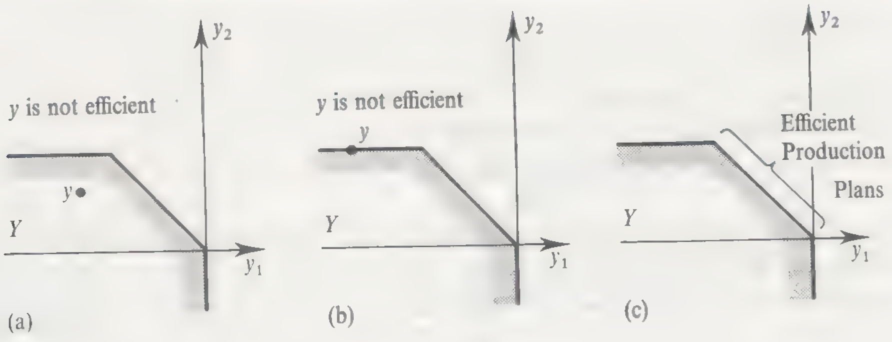


Figure 5.F.1
An efficient ... plan must be o... boundary of Y. ... not all points o... boundary of Y ar... efficient.
(a) An ineffici... production pla... interior of Y.
(b) An ineffici... production pla... boundary of Y.
(c) The set of ... production plans.

Definition 5.F.1: A production vector $y \in Y$ is *efficient* if there is no $y' \in Y$ such that $y' \geq y$ and $y' \neq y$.

In words, a production vector is efficient if there is no other feasible production vector that generates as much output as y using no additional inputs, and that actually produces more of some output or uses less of some input.

As we see in Figure 5.F.1, every efficient y must be on the boundary of Y, but the converse is not necessarily the case: There may be boundary points of Y that are not efficient.

We now show that the concept of efficiency is intimately related to that of supportability by profit maximization. This constitutes our first look at a topic that we explore in much more depth in Chapter 10 and especially in Chapter 16

Proposition 5.F.1 provides an elementary but important result. It is a version of the *first fundamental theorem of welfare economics.*

Proposition 5.F.1: If $y \in Y$ is profit maximizing for some $p \gg 0$, then y is efficient.

Proof: Suppose otherwise: That there is a $y' \in Y$ such that $y' \neq y$ and $y' \geq y$. Because $p \gg 0$, this implies that $p \cdot y' > p \cdot y$, contradicting the assumption that y is profit maximizing. ■

It is worth emphasizing that Proposition 5.F.1 is valid even if the production set is nonconvex. This is illustrated in Figure 5.F.2.

When combined with the aggregation results discussed in Section 5.E, Proposition 5.F.1 tells us that *if a collection of firms each independently maximizes profits with respect to the same fixed price vector $p \gg 0$, then the aggregate production is*

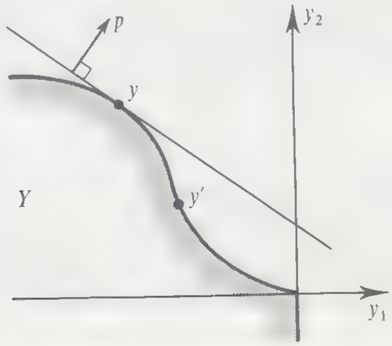


Figure 5.F.2
A profit-maximi... production plan ... $p \gg 0$) is efficient.

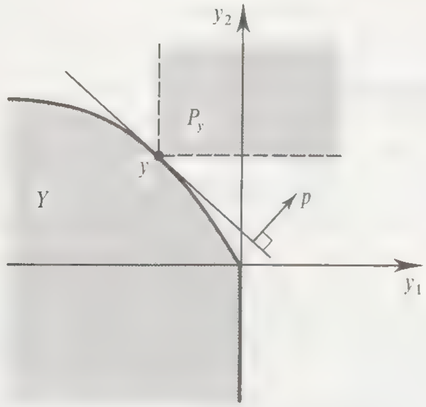


Figure 5.F.3 The use of the separating hyperplane theorem to prove Proposition 5.F.2: If Y is convex, every efficient $y \in Y$ is profit maximizing for some $p \geq 0$.

...ially *efficient*. That is, there is no other production plan for the economy as a whole that could produce more output using no additional inputs. This is in line with our conclusion in Section 5.E that, in the single-output case, the aggregate output level is produced at the lowest-possible cost when all firms maximize profits facing the same prices.

The need for strictly positive prices in Proposition 5.F.1 is unpleasant, but it cannot be dispensed with, as Exercise 5.F.1 asks you to demonstrate.

Exercise 5.F.1: Give an example of a $y \in Y$ that is profit maximizing for some $p \geq 0$ with $p \neq 0$ but that is also inefficient (i.e. not efficient).

A converse of Proposition 5.F.1 would assert that any efficient production vector is profit maximizing for *some* price system. However, a glance at the efficient production y' in Figure 5.F.2 shows that this cannot be true in general. Nevertheless, this converse does hold with the added assumption of convexity. Proposition 5.F.2, which is less elementary than Proposition 5.F.1, is a version of the so-called *second fundamental theorem of welfare economics.*

Proposition 5.F.2: Suppose that Y is convex. Then every efficient production $y \in Y$ is a profit-maximizing production for some nonzero price vector $p \geq 0$.[13]

Proof: This proof is an application of the separating hyperplane theorem for convex sets (see Section M.G of the Mathematical Appendix). Suppose that $y \in Y$ is efficient, and define the set $P_y = \{y' \in \mathbb{R}^L: y' \gg y\}$. The set P_y is depicted in Figure 5.F.3. It is convex, and because y is efficient, we have $Y \cap P_y = \varnothing$. We can therefore invoke the separating hyperplane theorem to establish that there is *some* $p \neq 0$ such that $p \cdot y' \geq p \cdot y''$ for every $y' \in P_y$ and $y'' \in Y$ (see Figure 5.F.3). Note, in particular, that this implies $p \cdot y' \geq p \cdot y$ for every $y' \gg y$. Therefore, we must have $p \geq 0$ because if $p_\ell < 0$ for some ℓ, then we would have $p \cdot y' < p \cdot y$ for some $y' \gg y$ with $y'_\ell - y_\ell$ sufficiently large.

Now take any $y'' \in Y$. Then $p \cdot y' \geq p \cdot y''$ for every $y' \in P_y$. Because y' can be chosen to be arbitrarily close to y, we conclude that $p \cdot y \geq p \cdot y''$ for any $y'' \in Y$; that is, y is profit maximizing for p. ■

13. As the proof makes clear, the result also applies to *weakly efficient* productions, that is, to productions such as y in Figure 5.F.1(b) where there is no $y' \in Y$ such that $y' \gg y$.

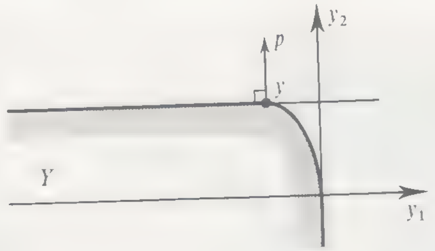


Figure 5.F.4
Proposition 5.C.2 cannot be extended to require $p \gg 0$.

The second part of Proposition 5.F.2 cannot be strengthened to read "$p \gg 0$." In Figure 5.F.4, for example, the production vector y is efficient, but it cannot be supported by any strictly positive price vector.

As an illustration of Proposition 5.F.2, consider a single-output, concave production function $f(z)$. Fix an input vector $\bar{z}$, and suppose that $f(\cdot)$ is differentiable at $\bar{z}$ and $\nabla f(\bar{z}) \gg 0$. Then the production plan that uses input vector $\bar{z}$ to produce output level $f(\bar{z})$ is efficient. Letting the price of output be 1, condition (5.C.2) tells us that the input price vector that makes this efficient production profit maximizing is precisely $w = \nabla f(\bar{z})$, the vector of marginal productivities.

5.G Remarks on the Objectives of the Firm

Although it is logical to take the assumption of preference maximization as a primitive concept for the theory of the consumer, the same cannot be said for the assumption of profit maximization by the firm. Why this objective rather than, say, the maximization of sales revenues or the size of the firm's labor force? The objectives of the firm assumed in our economic analysis should emerge from the objectives of those individuals who control it. Firms in the type of economies we consider are owned by individuals who, wearing another hat, are also consumers. A firm owned by a single individual has well-defined objectives: those of the owner. In this case, the only issue is whether this objective coincides with profit maximization. Whenever there is more than one owner, however, we have an added level of complexity. Indeed, we must either reconcile any conflicting objectives the owners may have or show that no conflict exists.

Fortunately, it is possible to resolve these issues and give a sound theoretical grounding to the objective of profit maximization. We shall now show that under reasonable assumptions this is the goal that all owners would agree upon.

Suppose that a firm with production set Y is owned by consumers. Ownership here simply means that each consumer $i = 1, \dots, I$ is entitled to a share $\theta_i \geq 0$ of profits, where $\sum_i \theta_i = 1$ (some of the θ_i's may equal zero). Thus, if the production decision is $y \in Y$, then a consumer i with utility function $u_i(\cdot)$ achieves the utility level

$$\begin{aligned} \underset{x_i \geq 0}{\text{Max}} \quad & u_i(x_i) \\ & \text{s.t. } p \cdot x_i \leq w_i + \theta_i p \cdot y, \end{aligned}$$

where w_i is consumer i's nonprofit wealth. Hence at fixed prices, higher profit increases consumer-owner i's overall wealth and expands her budget set, a desirable outcome. It follows that at any fixed price vector p, the consumer–owners *unanimously*

prefer that the firm implement a production plan $y' \in Y$ instead of $y \in Y$ whenever $p \cdot y' > p \cdot y$. Hence, we conclude that if we maintain the assumption of price-taking behavior, all owners would agree, whatever their utility functions, to instruct the manager of the firm to maximize profits.[14]

It is worth emphasizing three of the implicit assumptions in the previous reasoning: (i) prices are fixed and do not depend on the actions of the firm, (ii) profits are not uncertain, and (iii) managers can be controlled by owners. We comment on these assumptions very informally.

(i) If prices may depend on the production of the firm, the objective of the owners may depend on their tastes as consumers. Suppose, for example, that each consumer has no wealth from sources other than the firm ($w_i = 0$), that $L = 2$, and that the firm produces good 1 from good 2 with production function $f(\cdot)$. Also, normalize the price of good 2 to be 1, and suppose that the price of good 1, in terms of good 2, is $p(q)$ if output is q. If, for example, the preferences of the owners are such that they care only about the consumption of good 2, then they will unanimously want to solve $\text{Max}_{z \geq 0}\ p(f(z))f(z) - z$. This maximizes the amount of good 2 that they get to consume. On the other hand, if they want to consume only good 1, then they will wish to solve $\text{Max}_{z \geq 0}\ f(z) - [z/p(f(z))]$ because if they earn $p(f(z))f(z) - z$ units of good 2, then end up with $[p(f(z))f(z) - z]/p(f(z))$ units of good 1. But these two problems have different solutions. (Check the first-order conditions.) Moreover, as this suggests, if the owners differ in their tastes as consumers, then they will not agree about what they want the firm to do (Exercise 5.G.1 elaborates on this point.)

(ii) If the output of the firm is random, then it is crucial to distinguish whether the output is sold before or after the uncertainty is resolved. If the output is sold after the uncertainty is resolved (as in the case of agricultural products sold in spot markets after harvesting), then the argument for a unanimous desire for profit maximization breaks down. Because profit, and therefore derived wealth, are now uncertain, the risk attitudes and expectations of owners will influence their preferences with regard to production plans. For example, strong risk averters will prefer relatively less risky production plans than moderate risk averters.

On the other hand, if the output is sold before uncertainty is resolved (as in the case of agricultural products sold in futures markets before harvesting), then the risk is fully carried by the buyer. The profit of the firm is not uncertain, and the argument for unanimity in favor of profit maximization still holds. In effect, the firm can be thought of as producing a commodity that is sold before uncertainty is resolved in a market of the usual kind. (Further analysis of this issue would take us too far afield. We come back to it in Section 19.G after covering the foundations of decision theory under uncertainty in Chapter 6.)

(iii) It is plain that shareholders cannot usually exercise control directly. They need managers, who, naturally enough, have their own objectives. Especially if ownership is very diffuse, it is an important theoretical challenge to understand how and to what extent managers are, or can be, controlled by owners. Some relevant considerations are factors such as the degree of observability of managerial actions

14. In actuality, there are public firms and quasipublic organizations such as universities that do not have *owners* in the sense that private firms have shareholders. Their objectives may be different, and the current discussion does not apply to them.

and the stake of individual owners. [These issues will be touched on in Section 14.C (agency contracts as a mechanism of internal control) and in Section 19.G (stock markets as a mechanism of external control).]

APPENDIX A: THE LINEAR ACTIVITY MODEL

The saliency of the model of production with convexity and constant returns to scale technologies recommends that we examine it in some further detail.

Given a constant returns to scale technology Y, the *ray* generated (or spanned) by a vector $\bar{y} \in Y$ is the set $\{y \in Y: y = \alpha\bar{y}$ for some scalar $\alpha \geq 0\}$. We can think of a ray as representing a production *activity* that can be run at any *scale of operation.* That is, the production plan $\bar{y}$ can be scaled up or down by any factor $\alpha \geq 0$, generating, in this way, other possible production plans.

We focus here on a particular case of constant returns to scale technologies that lends itself to explicit computation and is therefore very important in applications. We assume that we are given as a primitive of our theory a list of *finitely many activities* (say M), each of which can be run at any scale of operation and any number of which can be run simultaneously. Denote the M activities, to be called the *elementary activities*, by $a_1 \in \mathbb{R}^L, \ldots, a_M \in \mathbb{R}^L$. Then, the production set is

$$Y = \{y \in \mathbb{R}^L: y = \sum_{m=1}^{M} \alpha_m a_m \text{ for some scalars } (\alpha_1, \ldots, \alpha_M) \geq 0\}.$$

The scalar α_m is called the *level of elementary activity m*; it measures the scale of operation of the mth activity. Geometrically, Y is a *polyhedral cone*, a set generated as the convex hull of a finite number of rays.

An activity of the form $(0, \ldots, 0, -1, 0, \ldots, 0)$, where -1 is in the ℓth place, is known as the *disposal activity* for good ℓ. Henceforth, we shall always assume that, in addition to the M listed elementary activities, the L disposal activities are also available. Figure 5.AA.1 illustrates a production set arising in the case where $L = 2$ and $M = 2$.

Given a price vector $p \in \mathbb{R}^L_+$, a profit-maximizing plan exists in Y if and only if $p \cdot a_m \leq 0$ for every m. To see this, note that if $p \cdot a_m < 0$, then the profit-maximizing level of activity m is $\alpha_m = 0$. If $p \cdot a_m = 0$, then any level of activity m generates zero profits. Finally, if $p \cdot a_m > 0$ for some m, then by making α_m arbitrarily large, we could generate arbitrarily large profits. Note that the presence of the disposal activities implies that we must have $p \in \mathbb{R}^L_+$ for a profit-maximizing plan to exist. If $p_\ell < 0$, then the ℓth disposal activity would generate strictly positive (hence, arbitrarily large) profits.

For any price vector p generating zero profits, let $A(p)$ denote the set of activities that generate exactly zero profits: $A(p) = \{a_m: p \cdot a_m = 0\}$. If $a_m \notin A(p)$, then $p \cdot a_m < 0$, and so activity m is not used at prices p. The profit-maximizing supply set $y(p)$ is therefore the convex cone generated by the activities in $A(p)$; that is, $y(p) = \{\sum_{a_m \in A(p)} \alpha_m a_m: \alpha_m \geq 0\}$. The set $y(p)$ is also illustrated in Figure 5.AA.1. In the figure, at price vector p, activity a_1 makes exactly zero profits, and activity a_2

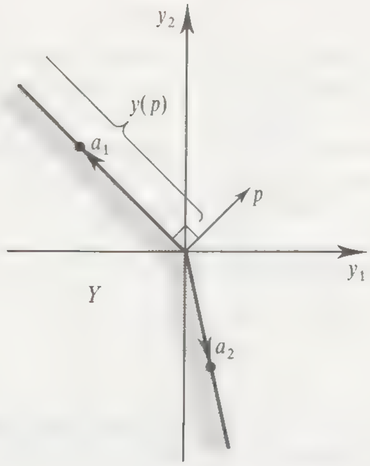


Figure 5.AA.1
A production set generated by two activities.

incurs a loss (if operated at all). Therefore, $A(p) = \{a_1\}$ and $y(p) = \{y: y = \alpha_1 a_1$ for any scalar $\alpha_1 \geq 0\}$, the ray spanned by activity a_1.

A significant result that we shall not prove is that for the linear activity model the converse of the efficiency Proposition 5.F.1 holds exactly; that is, we can strengthen Proposition 5.F.2 to say: *Every efficient $y \in Y$ is a profit-maximizing production for some $p \gg 0$.*

An important special case of the linear activity model is *Leontief's input–output model.* It is characterized by two additional features:

(i) There is one commodity, say the Lth, which is not produced by any activity. For this reason, we will call it the *primary factor.* In most applications of the Leontief model, the primary factor is labor.

(ii) Every elementary activity has at most a single positive entry. This is called the assumption of *no joint production.* Thus, it is as if every good except the primary factor is produced from a certain type of constant returns production function using the other goods and the primary factor as inputs.

The Leontief Input–Output Model with No Substitution Possibilities

The simplest Leontief model is one in which each producible good is produced by only one activity. In this case, it is natural to label the activity that produces good $\ell = 1, \ldots, L-1$ as $a_\ell = (a_{1\ell}, \ldots, a_{L\ell}) \in \mathbb{R}^L$. So the number of elementary activities M is equal to $L - 1$. As an example, in Figure 5.AA.2, for a case where $L = 3$, we represent the unit production isoquant [the set $\{(z_2, z_3): f(z_2, z_3) = 1\}$] for the implied production function of good 1. In the figure, the disposal activities for goods 2 and 3 are used to get rid of any excess of inputs. Because inputs must be used in fixed proportions (disposal aside), this special case is called a *Leontief model with no substitution possibilities.*

If we normalize the activity vectors so that $a_{\ell\ell} = 1$ for all $\ell = 1, \ldots, L-1$, then the vector $\alpha = (\alpha_1, \ldots, \alpha_{L-1}) \in \mathbb{R}^{L-1}$ of activity levels equals the vector of *gross* production of goods 1 through $L - 1$. To determine the levels of *net* production, it is convenient to denote by A the $(L-1) \times (L-1)$ matrix in which the ℓth column is

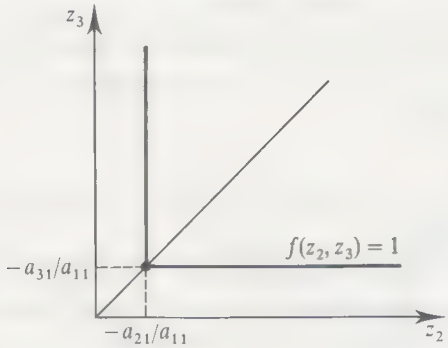


Figure 5.AA.2
Unit isoquant of production function for good 1 in the Leontief model with no substitution.

the negative of the activity vector a_ℓ except that its last entry has been deleted and entry $a_{\ell\ell}$ has been replaced by a zero (recall that entries $a_{k\ell}$ with $k \neq \ell$ are nonpositive):

$$A = \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1,L-1} \\ -a_{21} & 0 & \cdots & -a_{2,L-1} \\ \vdots & & \ddots & \\ -a_{L-1,1} & -a_{L-1,2} & \cdots & 0 \end{bmatrix}.$$

The matrix A is known as the *Leontief input–output matrix.* Its $k\ell$th entry, $-a_{k\ell} \geq 0$, measures how much of good k is needed to produce one unit of good ℓ. We also denote by $b \in \mathbb{R}^{L-1}$ the vector of primary factor requirements, $b = (-a_{L1}, \ldots, -a_{L,L-1})$. The vector $(I - A)\alpha$ then gives the *net* production levels of the $L - 1$ outputs when the activities are run at levels $\alpha = (\alpha_1, \ldots, \alpha_{L-1})$. To see this, recall that the activities are normalized so that the gross production levels of the $L - 1$ produced goods are exactly $\alpha = (\alpha_1, \ldots, \alpha_{L-1})$. On the other hand, $A\alpha$ gives the amounts of each of these goods that are used as inputs for other produced goods. The difference, $(I - A)\alpha$, is therefore the net production of goods $1, \ldots, L - 1$. In addition, the scalar $b \cdot \alpha$ gives the total use of the primary factor. In summary, with this notation, we can write the set of technologically feasible production vectors (assuming free disposal) as

$$Y = \left\{ y \colon y \leq \begin{bmatrix} I - A \\ -b \end{bmatrix} \alpha \text{ for some } \alpha \in \mathbb{R}^L_+ \right\}.$$

If $(I - A)\bar{\alpha} \gg 0$ for some $\bar{\alpha} \geq 0$, the input–output matrix A is said to be *productive.* That is, the input–output matrix A is productive if there is *some* production plan that can produce positive net amounts of the $L - 1$ outputs, provided only that there is a sufficient amount of primary input available.

A remarkable fact of Leontief input–output theory is the all-or-nothing property stated in Proposition 5.AA.1.

Proposition 5.AA.1: If A is productive, then for any nonnegative amounts of the $L - 1$ producible commodities $c \in \mathbb{R}^{L-1}_+$, there is a vector of activity levels $\alpha \geq 0$ such that $(I - A)\alpha = c$. That is, if A is productive, then it is possible to produce *any* nonnegative net amount of outputs (perhaps for purposes of final consumption), provided only that there is enough primary factor available.

Proof: We will show that if A is productive, then the inverse of the matrix $(I - A)$ exists and is nonnegative. This will give the result because we can then achieve net output levels $c \in \mathbb{R}^{L-1}_+$ by setting the (nonnegative) activity levels $\alpha = (I - A)^{-1}c$.

To prove the claim, we begin by establishing a matrix-algebra fact. We show that if A is productive, then the matrix $\sum_{n=0}^{N} A^n$, where A^n is the nth power of A, approaches a limit as $N \to \infty$. Because A has only nonnegative entries, every entry of $\sum_{n=0}^{N} A^n$ is *nondecreasing* with N. Therefore, to establish that $\sum_{n=0}^{N} A^n$ has a limit, it suffices to show that there is an upper bound for its entries. Since A is productive, there is an $\bar{\alpha}$ and $\bar{c} \gg 0$ such that $\bar{c} = (I - A)\bar{\alpha}$. If we premultiply both sides of this equality by $\sum_{n=0}^{N} A^n$, we get $(\sum_{n=0}^{N} A^n)\bar{c} = (I - A^{N+1})\bar{\alpha}$ (recall that $A^0 = I$). But $(I - A^{N+1})\bar{\alpha} \leq \bar{\alpha}$ because all elements of the matrix A^{N+1} are nonnegative. Therefore, $(\sum_{n=0}^{N} A^n)\bar{c} \leq \bar{\alpha}$. With $\bar{c} \gg 0$, this implies that no entry of $\sum_{n=0}^{N} A^n$ can exceed $[\text{Max}\,\{\bar{\alpha}_1, \ldots, \bar{\alpha}_{L-1}\}/\text{Min}\,\{\bar{c}_1, \ldots, \bar{c}_{L-1}\}]$, and so we have established the desired upper bound. We conclude, therefore, that $\sum_{n=0}^{\infty} A^n$ exists.

The fact that $\sum_{n=0}^{\infty} A^n$ exists must imply that $\lim_{N \to \infty} A^N = 0$. Thus, since $(\sum_{n=0}^{N} A^n)(I - A) = (I - A^{N+1})$ and $\lim_{N \to \infty} (I - A^{N+1}) = I$, it must be that $\sum_{n=0}^{\infty} A^n = (I - A)^{-1}$. (If A is a single number, this is precisely the high-school formula for adding up the terms of a geometric series.) The conclusion is that $(I - A)^{-1}$ exists and that all its entries are nonnegative. This establishes the result. ■

The focus on $\sum_{n=0}^{N} A^n$ in the proof of Proposition 5.AA.1 makes economic sense. Suppose we want to produce the vector of final consumptions $c \in \mathbb{R}^{L-1}_+$. How much total production will be needed? To produce final outputs $c = A^0 c$, we need to use as inputs the amounts $A(A^0 c) = Ac$ of produced goods. In turn, to produce these amounts requires that $A(Ac) = A^2 c$ of additional produced goods be used, and so on ad infinitum. The *total* amounts of goods required to be produced is therefore the limit of $(\sum_{n=0}^{N} A^n)c$ as $N \to \infty$. Thus, we can conclude that the vector $c \geq 0$ will be producible if and only if $\sum_{n=0}^{\infty} A^n$ is well defined (i.e., all its entries are finite).

Example 5.AA.1: Suppose that $L = 3$, and let $a_1 = (1, -1, -2)$ and $a_2 = (-\beta, 1, -4)$ for some constant $\beta \geq 0$. Activity levels $\alpha = (\alpha_1, \alpha_2)$ generate a positive net output of good 2 if $\alpha_2 > \alpha_1$; they generate a positive net output of good 1 if $\alpha_1 - \beta\alpha_2 > 0$. The input–output matrix A and the matrix $(I - A)^{-1}$ are

$$A = \begin{bmatrix} 0 & \beta \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad (I - A)^{-1} = \frac{1}{1 - \beta}\begin{bmatrix} 1 & \beta \\ 1 & 1 \end{bmatrix}.$$

Hence, matrix A is productive if and only if $\beta < 1$. Figure 5.AA.3(a) depicts a case where A is productive. The shaded region represents the vectors of net outputs that can be generated using the two activity vectors; note how the two activity vectors span all of $\mathbb{R}^2_+$. In contrast, in Figure 5.AA.3(b), the matrix A is not productive: no strictly positive vector of net outputs can be achieved by running the two activities at nonnegative scales. [Again, the shaded region represents those vectors that can be generated using the two activity vectors, here a set whose only intersection with $\mathbb{R}^2_+$ is the point $(0, 0)$]. Note also that the closer β is to the value 1, the larger the levels of activity required to produce any final vector of consumptions. ■

The Leontief Model with Substitution Possibilities

We now move to the consideration of the general Leontief model in which each good may have more than one activity capable of producing it. We shall see that the

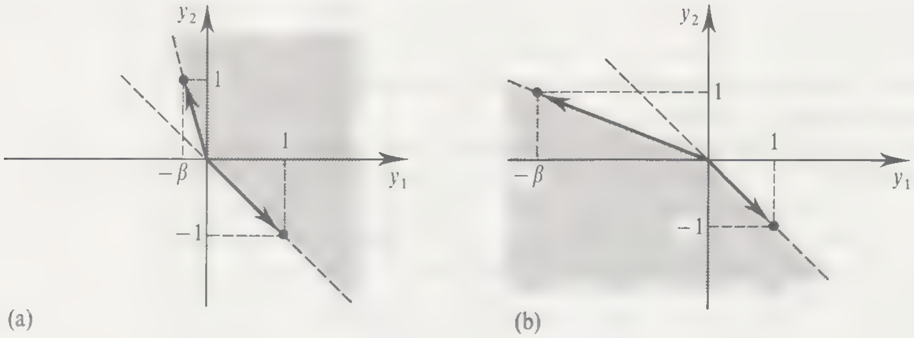


Figure 5.AA.3 Leontief model of Example 5.AA.1. (a) Productive ($\beta <$ (b) Unproductive ($\beta \geq 1$).

properties of the nonsubstitution model remain very relevant for the more general case where substitution is possible.

The first thing to observe is that the computation of the production function of a good, say good 1, now becomes a linear programming problem (see Section M.M of the Mathematical Appendix). Indeed, suppose that $a_1 \in \mathbb{R}^L, \dots, a_{M_1} \in \mathbb{R}^L$ is a list of M_1 elementary activities capable of producing good 1 and that we are given initial levels of goods $2, \dots, L$ equal to $z_2, \dots, z_L$. Then the maximal possible production of good 1 given these available inputs $f(z_2, \dots, z_L)$ is the solution to the problem

$$\begin{aligned} \underset{\alpha_1 \geq 0, \dots, \alpha_{M_1} \geq 0}{\text{Max}} \quad & \alpha_1 a_{11} + \cdots + \alpha_{M_1} a_{1M_1} \\ \text{s.t.} \quad & \textstyle\sum_{m=1}^{M_1} \alpha_m a_{\ell m} \geq -z_\ell \quad \text{for all } \ell = 2, \dots, L. \end{aligned}$$

We also know from linear programming theory that the $L-1$ dual variables $(\lambda_2, \dots, \lambda_L)$ of this problem (i.e., the multipliers associated with the $L-1$ constraints) can be interpreted as the marginal productivities of the $L-1$ inputs. More precisely, for any $\ell = 2, \dots, L$, we have $(\partial f/\partial z_\ell)^+ \leq \lambda_\ell \leq (\partial f/\partial z_\ell)^-$, where $(\partial f/\partial z_\ell)^+$ and $(\partial f/\partial z_\ell)^-$ are, respectively, the left-hand and right-hand ℓth partial derivatives of $f(\cdot)$ at $(z_2, \dots, z_L)$.

Figure 5.AA.4 illustrates the unit isoquant for the case in which good 1 can be produced using two other goods (goods 2 and 3) as inputs with two possible activities $a_1 = (1, -2, -1)$ and $a_2 = (1, -1, -2)$. If the ratio of inputs is either higher than 2 or lower than $\frac{1}{2}$, one of the disposal activities is used to eliminate any excess inputs.

For any vector $y \in \mathbb{R}^L$, it will be convenient to write $y = (y_{-L}, y_L)$, where $y_{-L} = (y_1, \dots, y_{L-1})$. We shall assume that our Leontief model is *productive* in the sense that there is a technologically feasible vector $y \in Y$ such that $y_{-L} \gg 0$.

A striking implication of the Leontief structure (constant returns, no joint products, single primary factor) is that we can associate with each good a *single optimal technique* (which could be a mixture of several of the elementary techniques corresponding to that good). What this means is that optimal techniques (one for each output) supporting efficient production vectors can be chosen independently of the particular output vector that is being produced (as long as the net output of every producible good is positive). Thus, although substitution is possible in principle, efficient production requires no substitution of techniques as desired final consumption levels change. This is the content of the celebrated *non-substitution theorem* (due to Samuelson [1951]).

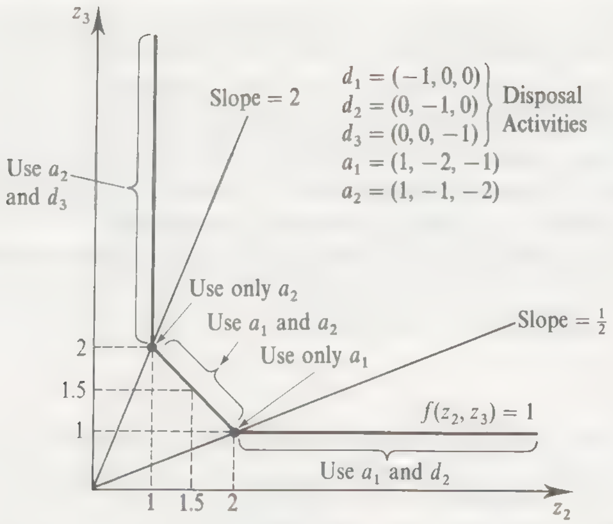


Figure 5.AA.4

Unit isoquant of production function of good 1, in the Leontief model with substitution.

Proposition 5.AA.2: (*The Nonsubstitution Theorem*) Consider a productive Leontief input–output model with $L-1$ producible goods and $M_\ell \geq 1$ elementary activities for the producible good $\ell = 1, \dots, L-1$. Then there exist $L-1$ activities $(a_1, \dots, a_{L-1})$, with a_ℓ possibly a nonnegative linear combination of the M_ℓ elementary activities for producing good ℓ, such that *all* efficient production vectors with $y_{-L} \gg 0$ can be generated with these $L-1$ activities.

Proof: Let $y \in Y$ be an efficient production vector with $y_{-L} \gg 0$. As a general matter, the vector y must be generated by a collection of $L-1$ activities $(a_1, \dots, a_{L-1})$ (some of these may be "mixtures" of the original activities) run at activity levels $(\alpha_1, \dots, \alpha_{L-1}) \gg 0$; that is, $y = \sum_{\ell=1}^{L-1} \alpha_\ell a_\ell$. We show that any efficient production plan y' with $y'_{-L} \gg 0$ can be achieved using the activities $(a_1, \dots, a_{L-1})$.

Since $y \in Y$ is efficient, there exists a $p \gg 0$ such that y is profit maximizing with respect to p (this is from Proposition 5.F.2, as strengthened for the linear activity model). From $p \cdot a_\ell \leq 0$ for all $\ell = 1, \dots, L-1$, $\alpha_\ell > 0$, and

$$0 = p \cdot y = p \cdot \left(\sum_{\ell=1}^{L-1} \alpha_\ell a_\ell \right) = \sum_{\ell=1}^{L-1} \alpha_\ell p \cdot a_\ell,$$

it follows that $p \cdot a_\ell = 0$ for all $\ell = 1, \dots, L-1$.

Consider now any other efficient production $y' \in Y$ with $y'_{-L} \gg 0$. We want to show that y' can be generated from the activities $(a_1, \dots, a_{L-1})$. Denote by A the input–output matrix associated with $(a_1, \dots, a_{L-1})$. Because $y_{-L} \gg 0$, it follows by assumption that A is productive. Therefore, by Proposition 5.AA.1, we know that there are activity levels $(\alpha''_1, \dots, \alpha''_{L-1})$ such that the production vector $y'' = \sum_{\ell=1}^{L-1} \alpha''_\ell a_\ell$ has $y''_{-L} = y'_{-L}$. Note that since $p \cdot a_\ell = 0$ for all $\ell = 1, \dots, L-1$, we must have $p \cdot y'' = 0$. Thus, y'' is profit maximizing for $p \gg 0$ (recall that the maximum profits for p are zero), and so it follows that y'' is efficient by Proposition 5.F.1. But then we have two production vectors, y' and y'', with $y'_{-L} = y''_{-L}$, and both are efficient. It must therefore be that $y''_L = y'_L$. Hence, we conclude that y' can be produced using only the activities $(a_1, \dots, a_{L-1})$, which is the desired result. ■

The nonsubstitution theorem depends critically on the presence of only one

primary factor. This makes sense. With more than one primary factor, the optimal choice of techniques should depend on the relative prices of these factors. In turn, it is logical to expect that these relative prices will not be independent of the composition of final demand (e.g., if demand moves from land-intensive goods toward labour-intensive goods, we would expect the price of labor relative to the price of land to increase). Nonetheless, it is worth mentioning that the nonsubstitution result remains valid as long as the prices of the primary factors do not change.

For further reading on the material discussed in this appendix see Gale (1960).

REFERENCES

Champsaur, P., and J.-C. Milleron (1983). *Advanced Exercises in Microeconomics.* Cambridge, Mass.: Harvard University Press.

Debreu, G. (1959). *Theory of Value.* New York: Wiley.

Gale, D. (1960). *The Theory of Linear Economic Models.* New York: McGraw-Hill.

Koopmans, T. (1957). *Three Essays on the State of Economic Science.* Essay 1. New York: McGraw-Hill.

McFadden, D. (1978). Cost, revenue and profit functions. In *Production economics: A dual approach to theory and applications,* edited by M. Fuss and D. McFadden. Amsterdam: North-Holland.

McKenzie, L. (1959). On the existence of a general equilibrium for a competitive market. *Econometrica* **27**: 54–71.

Samuelson, P. (1951). Abstract of a theorem concerning substitutability in open Leontief models. In *Activity analysis of production and allocation,* edited by T. Koopmans. New York: Wiley.

EXERCISES

5.B.1$^{\mathrm{A}}$ In text

5.B.2$^{\mathrm{A}}$ In text.

5.B.3$^{\mathrm{A}}$ In text.

5.B.4$^{\mathrm{B}}$ Suppose that Y is a production set, interpreted now as the technology of a single production unit. Denote by Y^+ the additive closure of Y, that is, the smallest production set that is additive and contains Y (in other words, Y^+ is the total production set if technology Y can be replicated an arbitrary number of times). Represent Y^+ for each of the examples of production sets depicted graphically in Section 5.B. In particular, note that for the typical decreasing returns technology of Figure 5.B.5(a), the additive closure Y^+ violates the closedness condition (ii). Discuss and compare with the case corresponding to Figure 5.B.5(b), where Y^+ is closed.

5.B.5$^{\mathrm{C}}$ Show that if Y is closed and convex, and $-\mathbb{R}^L_+ \subset Y$, then free disposal holds.

5.B.6$^{\mathrm{B}}$ There are three goods. Goods 1 and 2 are inputs. The third, with amounts denoted by q, is an output. Output can be produced by two techniques that can be operated simultaneously or separately. The techniques are not necessarily linear. The first (respectively, the second) technique uses only the first (respectively, the second) input. Thus, the first (respectively, the second) technique is completely specified by $\phi_1(q_1)$ [respectively, $\phi_2(q_2)$], the minimal amount of input one (respectively, two) *sufficient* to produce the amount of output q_1 (respectively, q_2). The two functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are increasing and $\phi_1(0) = \phi_2(0) = 0$.

(a) Describe the three-dimensional production set associated with these two techniques. Assume free disposal.

(b) Give sufficient conditions on $\phi_1(\cdot), \phi_2(\cdot)$ for the production set to display additivity.

(c) Suppose that the input prices are w_1 and w_2. Write the first-order necessary conditions for profit maximization and interpret. Under which conditions on $\phi_1(\cdot), \phi_2(\cdot)$ will the necessary conditions be sufficient?

(d) Show that if $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are strictly concave, then a cost-minimizing plan cannot involve the simultaneous use of the two techniques. Interpret the meaning of the concavity requirement, and draw isoquants in the two-dimensional space of input uses.

5.C.1^A In text.

5.C.2^A In text.

5.C.3^B Establish properties (viii) and (ix) of Proposition 5.C.2. [*Hint*: Property (viii) is easy; (ix) is more difficult. Try the one-input case first.]

5.C.4^A Establish properties (i) to (vii) of Proposition 5.C.2 for the case in which there are multiple outputs.

5.C.5^A Argue that for property (iii) of Proposition 5.C.2 to hold, it suffices that $f(\cdot)$ be quasiconcave. Show that quasiconcavity of $f(\cdot)$ is compatible with increasing returns.

5.C.6^C Suppose $f(z)$ is a concave production function with $L-1$ inputs $(z_1, \ldots, z_{L-1})$. Suppose also that $\partial f(z)/\partial z_\ell \geq 0$ for all ℓ and $z \geq 0$ and that the matrix $D^2 f(z)$ is negative definite at all z. Use the firm's first-order conditions and the implicit function theorem to prove the following statements:

(a) An increase in the output price always increases the profit-maximizing level of output.

(b) An increase in output price increases the demand for *some* input.

(c) An increase in the price of an input leads to a reduction in the demand for the input.

5.C.7^C A price-taking firm producing a single product according to the technology $q = f(z_1, \ldots, z_{L-1})$ faces prices p for its output and $w_1, \ldots, w_{L-1}$ for each of its inputs. Assume that $f(\cdot)$ is strictly concave and increasing, and that $\partial^2 f(z)/\partial z_\ell\, \partial z_k < 0$ for all $\ell \neq k$. Show that for all $\ell = 1, \ldots, L-1$, the factor demand functions $z_\ell(p, w)$ satisfy $\partial z_\ell(p, w)/\partial p > 0$ and $\partial z_\ell(p, w)/\partial w_k < 0$ for all $k \neq \ell$.

5.C.8^B Alpha Incorporated (AI) produces a single output q from two inputs z_1 and z_2. You are assigned to determine AI's technology. You are given 100 monthly observations. Two of these monthly observations are shown in the following table:

	Input prices		*Input levels*		*Output price*	*Output level*
Month	w_1	w_2	z_1	z_2	p	q
3	3	1	40	50	4	60
95	2	2	55	40	4	60

In light of these two monthly observations, what problem will you encounter in trying to accomplish your task?

5.C.9^A Derive the profit function $\pi(p)$ and supply function (or correspondence) $y(p)$ for the single-output technologies whose production functions $f(z)$ are given by

(a) $f(z) = \sqrt{z_1 + z_2}$.

(b) $f(z) = \sqrt{\text{Min}\,\{z_1, z_2\}}$.

(c) $f(z) = (z_1^\rho + z_2^\rho)^{1/\rho}$, for $\rho \leq 1$.

5.C.10[A] Derive the cost function $c(w, q)$ and conditional factor demand functions (or correspondences) $z(w, q)$ for each of the following single-output constant return technologies with production functions given by

(a) $f(z) = z_1 + z_2$ (perfect substitutable inputs)

(b) $f(z) = \text{Min}\,\{z_1, z_2\}$ (Leontief technology)

(c) $f(z) = (z_1^\rho + z_2^\rho)^{1/\rho}$, $\rho \leq 1$ (constant elasticity of substitution technology)

5.C.11[A] Show that $\partial z_\ell(w, q)/\partial q > 0$ if and only if marginal cost at q is increasing in w_ℓ.

5.C.12[A] We saw at the end of Section 5.B that any convex Y can be viewed as the section of a constant returns technology $Y' \subset \mathbb{R}^{L+1}$, where the $L + 1$ coordinate is fixed at the level -1. Show that if $y \in Y$ is profit maximizing at prices p then $(y, -1) \in Y'$ is profit maximizing at $(p, \pi(p))$, that is, profits emerge as the price of the implicit fixed input. The converse is also true: If $(y, -1) \in Y'$ is profit maximizing at prices (p, p_{L+1}), then $y \in Y$ is profit maximizing at p and the profit is p_{L+1}.

5.C.13[B] A price-taking firm produces output q from inputs z_1 and z_2 according to a differentiable concave production function $f(z_1, z_2)$. The price of its output is $p > 0$, and the prices of its inputs are $(w_1, w_2) \gg 0$. However, there are two unusual things about this firm. First, rather than maximizing profit, the firm maximizes revenue (the manager wants her firm to have bigger dollar sales than any other). Second, the firm is cash constrained. In particular, it has only C dollars on hand before production and, as a result, its total expenditures on inputs cannot exceed C.

Suppose one of your econometrician friends tells you that she has used repeated observations of the firm's revenues under various output prices, input prices, and levels of the financial constraint and has determined that the firm's revenue level R can be expressed as the following function of the variables (p, w_1, w_2, C):

$$R(p, w_1, w_2, C) = p[\gamma + \ln C - \alpha \ln w_1 - (1 - \alpha) \ln w_2].$$

(γ and α are scalars whose values she tells you.) What is the firm's use of input z_1 when prices are (p, w_1, w_2) and it has C dollars of cash on hand?

5.D.1[A] In text.

5.D.2[A] In text.

5.D.3[B] Suppose that a firm can produce good L from $L - 1$ factor inputs ($L > 2$). Factor prices are $w \in \mathbb{R}^{L-1}$ and the price of output is p. The firm's differentiable cost function is $c(w, q)$. Assume that this function is strictly convex in q. However, although $c(w, q)$ is the cost function when all factors can be freely adjusted, factor 1 cannot be adjusted in the short run.

Suppose that the firm is initially at a point where it is producing its long-run profit-maximizing output level of good L given prices w and p, $q(w, p)$ [i.e., the level that is optimal under the long-run cost conditions described by $c(w, q)$], and that all inputs are optimally adjusted [i.e., $z_\ell = z_\ell(w, q(w, p))$ for all $\ell = 1, \ldots, L - 1$, where $z_\ell(\cdot, \cdot)$ is the long-run input demand function]. Show that the firm's profit-maximizing output response to a marginal increase in the price of good L is larger in the long run than in the short run. [*Hint*: Define a short-run cost function $c_s(w, q \mid z_1)$ that gives the minimized costs of producing output level q given that input 1 is fixed at level z_1.]

5.D.4^{B} Consider a firm that has a distinct set of inputs and outputs. The firm produces M outputs; let $q = (q_1, \ldots, q_M)$ denote a vector of its output levels. Holding factor prices fixed, $C(q_1, \ldots, q_M)$ is the firm's cost function. We say that $C(\cdot)$ is *subadditive* if for all $(q_1, \ldots, q_M)$, there is no way to break up the production of amounts $(q_1, \ldots, q_M)$ among several firms, each with cost function $C(\cdot)$, and lower the costs of production. That is, there is no set of, say, J firms and collection of production vectors $\{q_j = (q_{1j}, \ldots, q_{Mj})\}_{j=1}^{J}$ such that $\sum_j q_j = q$ and $\sum_j C(q_j) < C(q)$. When $C(\cdot)$ is subadditive, it is usual to say that the industry is a *natural monopoly* because production is cheapest when it is done by only one firm.

(a) Consider the single-output case, $M = 1$. Show that if $C(\cdot)$ exhibits decreasing average costs, then $C(\cdot)$ is subadditive.

(b) Now consider the multiple-output case, $M > 1$. Show by example that the following multiple-output extension of the decreasing average cost assumption is *not* sufficient for $C(\cdot)$ to be subadditive:

$C(\cdot)$ exhibits *decreasing ray average cost* if for any $q \in \mathbb{R}_{+}^{M}$,

$$C(q) > C(kq)/k \text{ for all } k > 1.$$

(c) (Harder) Prove that, if $C(\cdot)$ exhibits decreasing ray average cost *and* is quasiconvex, then $C(\cdot)$ is subadditive. [Assume that $C(\cdot)$ is continuous, increasing, and satisfies $C(0) = 0$.]

5.D.5^{B} Suppose there are two goods: an input z and an output q. The production function is $q = f(z)$. We assume that $f(\cdot)$ exhibits increasing returns to scale.

(a) Assume that $f(\cdot)$ is differentiable. Do the increasing returns of $f(\cdot)$ imply that the average product is necessarily nondecreasing in input? What about the marginal product?

(b) Suppose there is a representative consumer with the utility function $u(q) - z$ (the negative sign indicates that the input is taken away from the consumer). Suppose that $\bar{q} = f(\bar{z})$ is a production plan that maximizes the representative consumer utility. Argue, either mathematically or economically (disregard boundary solutions), that the equality of marginal utility and marginal cost is a necessary condition for this maximization problem.

(c) Assume the existence of a representative consumer as in (b). "The equality of marginal cost and marginal utility is a sufficient condition for the optimality of a production plan." Right or wrong? Discuss.

5.E.1^{A} Assuming that every $\pi_j(\cdot)$ is differentiable and that you already know that $\pi^*(p) = \sum_{j=1}^{J} \pi_j(p)$, give a proof of $y^*(p) = \sum_{j=1}^{J} y_j(p)$ using differentiability techniques.

5.E.2^{A} Verify that Proposition 5.E.1 and its interpretation do not depend on any convexity hypothesis on the sets $Y_1, \ldots, Y_J$.

5.E.3^{B} Assuming that the sets $Y_1, \ldots, Y_J$ are convex and satisfy the free disposal property, and that $\sum_{j=1}^{J} Y_j$ is closed, show that the latter set equals $\{y\colon p \cdot y \leq \sum_{j=1}^{J} \pi_j(p) \text{ for all } p \gg 0\}$.

5.E.4^{B} One output is produced from two inputs. There are many technologies. Every technology can produce up to one unit of output (but no more) with fixed and proportional input requirements z_1 and z_2. So a technology is characterized by $z = (z_1, z_2)$, and we can describe the population of technologies by a density function $g(z_1, z_2)$. Take this density to be uniform on the square $[0, 10] \times [0, 10]$.

(a) Given the input prices $w = (w_1, w_2)$, solve the profit maximization problem of a firm with characteristics z. The output price is 1.

(b) More generally, find the profit function $\pi(w_1, w_2, 1)$ for

$$w_1 \geq \frac{1}{10} \quad \text{and} \quad w_2 \geq \frac{1}{10}.$$

(c) Compute the aggregate input demand function. Ideally, do that directly, and check that the answer is correct by using your finding in (b); this way you also verify (b).

(d) What can you say about the aggregate production function? If you were to assume that the profit function derived in (b) is valid for $w_1 \geq 0$ and $w_2 \geq 0$, what would the underlying aggregate production function be?

5.E.5^A (M. Weitzman) Suppose that there are J single-output plants. Plant j's average cost is $AC_j(q_j) = \alpha + \beta_j q_j$ for $q_j \geq 0$. Note that the coefficient α is the same for all plants but that the coefficient β_j may differ from plant to plant. Consider the problem of determining the cost-minimizing aggregate production plan for producing a total output of q, where $q < (\alpha/\text{Max}_j |\beta_j|)$.

(a) If $\beta_j > 0$ for all j, how should output be allocated among the J plants?

(b) If $\beta_j < 0$ for all j, how should output be allocated among the J plants?

(c) What if $\beta_j > 0$ for some plants and $\beta_j < 0$ for others?

5.F.1^A In text.

5.G.1^B Let $f(z)$ be a single-input, single-output production function. Suppose that owners have quasilinear utilities with the firm's input as the numeraire.

(a) Show that a necessary condition for consumer–owners to unanimously agree to a production plan z is that consumption shares among owners at prices $p(z)$ coincide with ownership shares.

(b) Suppose that ownership shares are identical. Comment on the conflicting instructions to managers and how they depend on the consumer–owners' tastes for output.

(c) With identical preferences and ownership shares, argue that owners will unanimously agree to maximize profits in terms of input. (Recall that we are assuming preferences are quasilinear with respect to input; hence, the numeraire is intrinsically determined.)

5.AA.1^A Compute the cost function $c(w, 1)$ and the input demand $z(w, 1)$ for the production function in Figure 5.AA.4. Verify that whenever $z(w, 1)$ is single-valued, we have $z(w, 1) = \nabla_w c(w, 1)$.

5.AA.2^B Consider a Leontief input–output model with no substitution. Assume that the input matrix A is productive and that the vector of primary factor requirements b is strictly positive.

(a) Show that for any $\alpha \geq 0$, the production plan

$$y = \begin{bmatrix} I - A \\ -b \end{bmatrix} \alpha.$$

is efficient.

(b) Fixing the price of the primary factor to equal 1, show that any production plan with $\alpha \gg 0$ is profit maximizing at a unique vector of prices.

(c) Show that the prices obtained in (b) have the interpretation of amounts of the primary factor directly or indirectly embodied in the production of one unit of the different goods.

(d) (Harder) Suppose that A corresponds to the techniques singled out by the nonsubstitution theorem for a model that, in principle, admits substitution. Show that every component of the price vector obtained from A in (c) is less than or equal to the corresponding component of the price vector obtained from any other selection of techniques.

5.AA.3$^{\text{B}}$ There are two produced goods and labor. The input–output matrix is

$$A = \begin{bmatrix} 0 & 1 \\ \alpha & 0 \end{bmatrix}.$$

Here $a_{\ell k}$ is the amount of good ℓ required to produce one unit of good k.

(a) Let $\alpha = \frac{1}{2}$, and suppose that the labor coefficients vector is

$$b = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

where b_1 (respectively, b_2) is the amount of labor required to produce one unit of good 1 (respectively, good 2). Represent graphically the production possibility set (i.e., the locus of possible productions) for the two goods if the total availability of labor is 10.

(b) For the values of α and b in (a), compute equilibrium prices p_1, p_2 (normalize the wage as equal 1) from the profit maximization conditions (assume positive production of the two goods).

(c) For the values of α and b in (a), compute the amount of labor directly or indirectly incorporated into the production of one net (i.e., available for consumption) unit of good 1. How does this amount relate to your answer in **(b)**?

(d) Suppose there is a second technique to produce good 2. To

$$\begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad b_2 = 2$$

we now add

$$\begin{bmatrix} a'_{12} \\ a'_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}, \qquad b'_2 = \beta.$$

Taking the two techniques into account, represent graphically the locus of amounts of good 1 and of labor necessary to produce one unit of good 2. (Assume free disposal.)

(e) In the context of **(d)**, what does the nonsubstitution theorem say? Determine the value of β at which there is a switch of optimal techniques.

5.AA.4$^{\text{B}}$ Consider the following linear activity model:

$$\begin{aligned}
a_1 &= (1, -1, \quad 0, \quad 0) \\
a_2 &= (0, -1, \quad 1, \quad 0) \\
a_3 &= (0, \quad 0, -1, \quad 1) \\
a_4 &= (2, \quad 0, \quad 0, -1)
\end{aligned}$$

(a) For each of the following input–output vectors, check whether they belong or do not belong to the aggregate production set. Justify your answers:

$$\begin{aligned}
y_1 &= (6, \quad 0, \quad 0, -2) \\
y_2 &= (5, -3, \quad 0, -1) \\
y_3 &= (6, -3, \quad 0, \quad 0) \\
y_4 &= (0, -4, \quad 0, \quad 4) \\
y_5 &= (0, -3, \quad 4, \quad 0)
\end{aligned}$$

(b) The input–output vector $y = (0, -5, 5, 0)$ is efficient. Prove this by finding a $p \gg 0$ for which y is profit-maximizing.

(c) The input–output vector $y = (1, -1, 0, 0)$ is feasible, but it is not efficient. Why?

5.AA.5$^{\text{B}}$ [This exercise was inspired by an exercise of Champsaur and Milleron (1983).] There are four commodities indexed by $\ell = 1, 2, 3, 4$. The technology of a firm is described by eight

elementary activities a_m, $m = 1, \ldots, 8$. With the usual sign convention, the numerical values of these activities are

$$a_1 = (-3, -6, 4, 0)$$
$$a_2 = (-7, -9, 3, 2)$$
$$a_3 = (-1, -2, 3, -1)$$
$$a_4 = (-8, -13, 3, 1)$$
$$a_5 = (-11, -19, 12, 0)$$
$$a_6 = (-4, -3, -2, 5)$$
$$a_7 = (-8, -5, 0, 10)$$
$$a_8 = (-2, -4, 5, 2)$$

It is assumed that any activity can be operated at any nonnegative level $\alpha_m \geq 0$ and that all activities can operate simultaneously at any scale (i.e., for any $\alpha_m \geq 0$, $m = 1, \ldots, 8$, the production $\sum_m \alpha_m a_m$ is feasible).

(a) Define the corresponding production set Y, and show that it is convex.

(b) Verify the no-free-lunch property.

(c) Verify that Y does *not* satisfy the free-disposal property. The free-disposal property would be satisfied if we added new elementary activities to our list. How would you choose them (given specific numerical values)?

(d) Show by direct comparison of a_1 with a_5, a_2 with a_4, a_3 with a_8, and a_6 with a_7 that four of the elementary activities are not efficient.

(e) Show that a_1 and a_2 are inefficient be exhibiting two positive linear combinations of a_3 and a_7 that dominate a_1 and a_2, respectively.

(f) Could you venture a complete description of the set of efficient production vectors?

(g) Suppose that the amounts of the four goods available as initial resources to the firm are

$$s_1 = 480, \qquad s_2 = 300, \qquad s_3 = 0, \qquad s_4 = 0.$$

Subject to those limitations on the net use of resources, the firm is interested in maximizing the net production of the third good. How would you set up the problem as a linear program?

(h) By using all the insights you have gained on the set of efficient production vectors, can you solve the optimization problem in **(g)**? [*Hint*: It can be done graphically.]

CHAPTER

6

Choice Under Uncertainty

6.A Introduction

In previous chapters, we studied choices that result in perfectly certain outcomes. In reality, however, many important economic decisions involve an element of risk. Although it is formally possible to analyze these situations using the general theory of choice developed in Chapter 1, there is good reason to develop a more specialized theory: Uncertain alternatives have a structure that we can use to restrict the preferences that "rational" individuals may hold. Taking advantage of this structure allows us to derive stronger implications than those based solely on the framework of Chapter 1.

In Section 6.B, we begin our study of choice under uncertainty by considering a setting in which alternatives with uncertain outcomes are describable by means of objectively known probabilities defined on an abstract set of possible outcomes. These representations of risky alternatives are called *lotteries*. In the spirit of Chapter 1, we assume that the decision maker has a rational preference relation over these lotteries. We then proceed to derive the *expected utility theorem*, a result of central importance. This theorem says that under certain conditions, we can represent preferences by an extremely convenient type of utility function, one that possesses what is called the *expected utility form*. The key assumption leading to this result is the *independence axiom*, which we discuss extensively.

In the remaining sections, we focus on the special case in which the outcome of a risky choice is an amount of money (or any other one-dimensional measure of consumption). This case underlies much of finance and portfolio theory, as well as substantial areas of applied economics.

In Section 6.C, we present the concept of *risk aversion* and discuss its measurement. We then study the comparison of risk aversions both across different individuals and across different levels of an individual's wealth.

Section 6.D is concerned with the comparison of alternative distributions of monetary returns. We ask when one distribution of monetary returns can unambiguously be said to be "better" than another, and also when one distribution can be said to be "more risky than" another. These comparisons lead, respectively, to the concepts of *first-order* and *second-order stochastic dominance*.

In Section 6.E, we extend the basic theory by allowing utility to depend on *states of nature* underlying the uncertainty as well as on the monetary payoffs. In the process, we develop a framework for modeling uncertainty in terms of these underlying states. This framework is often of great analytical convenience, and we use it extensively later in this book.

In Section 6.F, we consider briefly the theory of *subjective probability*. The assumption that uncertain prospects are offered to us with known objective probabilities, which we use in Section 6.B to derive the expected utility theorem, is rarely descriptive of reality. The subjective probability framework offers a way of modeling choice under uncertainty in which the probabilities of different risky alternatives are not given to the decision maker in any objective fashion. Yet, as we shall see, the theory of subjective probability offers something of a rescue for our earlier objective probability approach.

For further reading on these topics, see Kreps (1988) and Machina (1987). Diamond and Rothschild (1978) is an excellent sourcebook for original articles.

6.B Expected Utility Theory

We begin this section by developing a formal apparatus for modeling risk. We then apply this framework to the study of preferences over risky alternatives and to establish the important expected utility theorem.

Description of Risky Alternatives

Let us imagine that a decision maker faces a choice among a number of risky alternatives. Each risky alternative may result in one of a number of possible *outcomes*, but which outcome will actually occur is uncertain at the time that he must make his choice.

Formally, we denote the set of all possible outcomes by C.[1] These outcomes could take many forms. They could, for example, be consumption bundles. In this case, $C = X$, the decision maker's consumption set. Alternatively, the outcomes might take the simpler form of monetary payoffs. This case will, in fact, be our leading example later in this chapter. Here, however, we treat C as an abstract set and therefore allow for very general outcomes.

To avoid some technicalities, we assume in this section that the number of possible outcomes in C is finite, and we index these outcomes by $n = 1, \ldots, N$.

Throughout this and the next several sections, we assume that the probabilities of the various outcomes arising from any chosen alternative are *objectively known*. For example, the risky alternatives might be monetary gambles on the spin of an unbiased roulette wheel.

The basic building block of the theory is the concept of a *lottery*, a formal device that is used to represent risky alternatives.

Definition 6.B.1: A *simple lottery* L is a list $L = (p_1, \ldots, p_N)$ with $p_n \geq 0$ for all n and $\sum_n p_n = 1$, where p_n is interpreted as the probability of outcome n occurring.

1. It is also common, following Savage (1954), to refer to the elements of C as *consequences*.

n: OUTCOMES

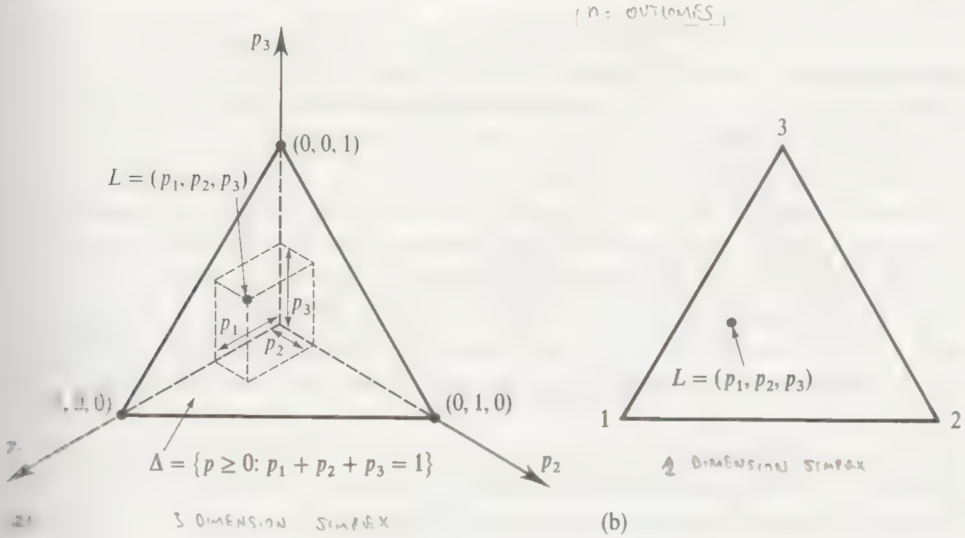


Figure 6.B.1
Representations of the simplex when $N = 3$.
(a) Three-dimensional representation.
(b) Two-dimensional representation.

A simple lottery can be represented geometrically as a point in the $(N-1)$ dimensional simplex, $\Delta = \{p \in \mathbb{R}^N_+ : p_1 + \cdots + p_N = 1\}$. Figure 6.B.1(a) depicts this simplex for the case in which $N = 3$. Each vertex of the simplex stands for the degenerate lottery where one outcome is certain and the other two outcomes have probability zero. Each point in the simplex represents a lottery over the three outcomes. When $N = 3$, it is convenient to depict the simplex in two dimensions, as in Figure 6.B.1(b), where it takes the form of an equilateral triangle.[2]

In a simple lottery, the outcomes that may result are certain. A more general variant of a lottery, known as a *compound lottery*, allows the outcomes of a lottery themselves to be simple lotteries.[3]

COMPOUND LOTTERY

Definition 6.B.2: Given K simple lotteries $L_k = (p_1^k, \ldots, p_N^k)$, $k = 1, \ldots, K$, and probabilities $\alpha_k \geq 0$ with $\sum_k \alpha_k = 1$, the *compound lottery* $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$ is the risky alternative that yields the simple lottery L_k with probability α_k for $k = 1, \ldots, K$.

For any compound lottery $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$, we can calculate a corresponding *reduced lottery* as the simple lottery $L = (p_1, \ldots, p_N)$ that generates the same ultimate distribution over outcomes. The value of each p_n is obtained by multiplying the probability that each lottery L_k arises, α_k, by the probability p_n^k that outcome n arises in lottery L_k, and then adding over k. That is, the probability of outcome n in the reduced lottery is

$$p_n = \alpha_1 p_n^1 + \cdots + \alpha_K p_n^K,$$

N_?B

2. Recall that equilateral triangles have the property that the sum of the perpendiculars from any point to the three sides is equal to the altitude of the triangle. It is therefore common to depict the simplex when $N = 3$ as an equilateral triangle with altitude equal to 1 because by doing so, we have the convenient geometric property that the probability p_n of outcome n in the lottery associated with some point in this simplex is equal to the length of the perpendicular from this point to the side opposite the vertex labeled n.

3. We could also define compound lotteries with more than two stages. We do not do so because we will not need them in this chapter. The principles involved, however, are the same.

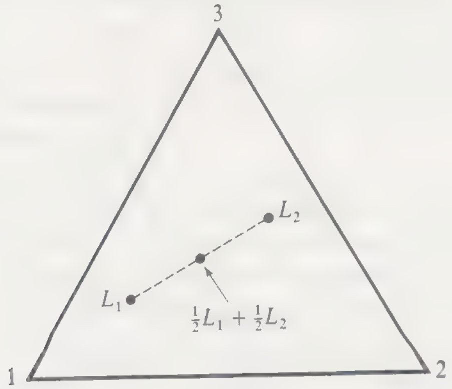


Figure 6.B.2
The
a co

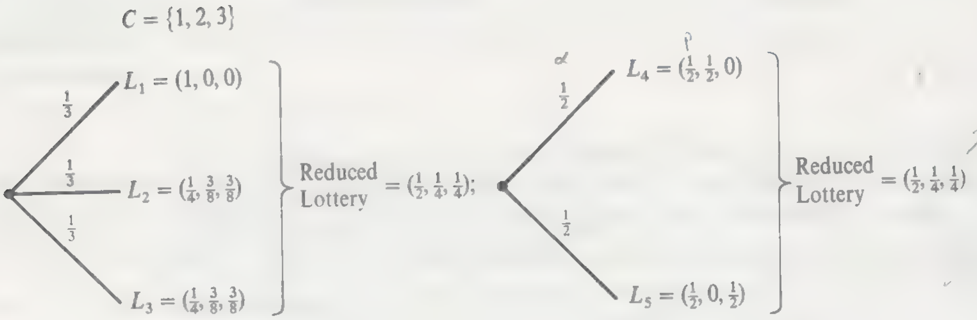


Figure 6.B.3
Two
lotteries
reduced

for $n = 1, \dots, N$.[4] Therefore, the reduced lottery L of any compound lottery $(L_1, \dots, L_K; \alpha_1, \dots, \alpha_K)$ can be obtained by vector addition:

$$L = \alpha_1 L_1 + \cdots + \alpha_K L_K \in \Delta.$$

In Figure 6.B.2, two simple lotteries L_1 and L_2 are depicted in the simplex Δ. Also depicted is the reduced lottery $\frac{1}{2}L_1 + \frac{1}{2}L_2$ for the compound lottery $(L_1, L_2; \frac{1}{2}, \frac{1}{2})$ that yields either L_1 or L_2 with a probability of $\frac{1}{2}$ each. This reduced lottery lies at the midpoint of the line segment connecting L_1 and L_2. The linear structure of the space of lotteries is central to the theory of choice under uncertainty, and we exploit it extensively in what follows.

Preferences over Lotteries

Having developed a way to model risky alternatives, we now study the decision maker's preferences over them. The theoretical analysis to follow rest on a basic *consequentialist* premise: We assume that for any risky alternative, only the reduced lottery over final outcomes is of relevance to the decision maker. Whether the probabilities of various outcomes arise as a result of a simple lottery or of a more complex compound lottery has no significance. Figure 6.B.3 exhibits two different compound lotteries that yield the same reduced lottery. Our consequentialist hypothesis requires that the decision maker view these two lotteries as equivalent.

---

4. Note that $\sum_n p_n = \sum_k \alpha_k (\sum_n p_n^k) = \sum_k \alpha_k = 1$.

We now pose the decision maker's choice problem in the general framework developed in Chapter 1 (see Section 1.B). In accordance with our consequentialist premise, we take the set of alternatives, denoted here by $\mathscr{L}$, to be *the set of all simple lotteries over the set of outcomes C*. We next assume that the decision maker has a rational preference relation $\succsim$ on $\mathscr{L}$, a complete and transitive relation allowing comparison of any pair of simple lotteries. It should be emphasized that, if anything, the rationality assumption is stronger here than in the theory of choice under certainty discussed in Chapter 1. The more complex the alternatives, the heavier the burden carried by the rationality postulates. In fact, their realism in an uncertainty context has been much debated. However, because we want to concentrate on the properties that are specific to uncertainty, we do not question the rationality assumption further here.

We next introduce two additional assumptions about the decision maker's preferences over lotteries. The most important and controversial is the *independence axiom*. The first, however, is a continuity axiom similar to the one discussed in Section 3.C.

Definition 6.B.3: The preference relation $\succsim$ on the space of simple lotteries $\mathscr{L}$ is *continuous* if for any $L, L', L'' \in \mathscr{L}$, the sets

CONTINUITY

$$\{\alpha \in [0,1]: \alpha L + (1-\alpha)L' \succsim L''\} \subset [0,1]$$

and

$$\{\alpha \in [0,1]: L'' \succsim \alpha L + (1-\alpha)L'\} \subset [0,1]$$

are closed.

In words, continuity means that small changes in probabilities do not change the nature of the ordering between two lotteries. For example, if a "beautiful and uneventful trip by car" is preferred to "staying home," then a mixture of the outcome "beautiful and uneventful trip by car" with a sufficiently small but positive probability of "death by car accident" is still preferred to "staying home." Continuity therefore rules out the case where the decision maker has lexicographic ("safety first") preferences for alternatives with a zero probability of some outcome (in this case, "death by car accident").

As in Chapter 3, the continuity axiom implies the existence of a utility function representing $\succsim$, a function $U: \mathscr{L} \to \mathbb{R}$ such that $L \succsim L'$ if and only if $U(L) \geq U(L')$. The second assumption, the independence axiom, will allow us to impose considerably more structure on $U(\cdot)$.[5]

Definition 6.B.4: The preference relation $\succsim$ on the space of simple lotteries $\mathscr{L}$ satisfies the *independence axiom* if for all $L, L', L'' \in \mathscr{L}$ and $\alpha \in (0,1)$ we have

INDEPENDENCE AXM

$$L \succsim L' \quad \text{if and only if} \quad \alpha L + (1-\alpha)L'' \succsim \alpha L' + (1-\alpha)L''.$$

In other words, if we mix each of two lotteries with a third one, then the preference ordering of the two resulting mixtures does not depend on (is *independent* of) the particular third lottery used.

[5] The independence axiom was first proposed by von Neumann and Morgenstern (1944) as an incidental result in the theory of games.

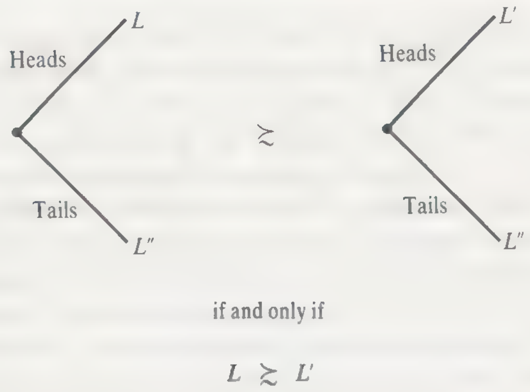


Figure 6.B.4 The independence axiom.

Suppose, for example, that $L \succsim L'$ and $\alpha = \frac{1}{2}$. Then $\frac{1}{2}L + \frac{1}{2}L''$ can be thought of as the compound lottery arising from a coin toss in which the decision maker gets L if heads comes up and L'' if tails does. Similarly, $\frac{1}{2}L' + \frac{1}{2}L''$ would be the coin toss where heads results in L' and tails results in L'' (see Figure 6.B.4). Note that conditional on heads, lottery $\frac{1}{2}L + \frac{1}{2}L''$ is at least as good as lottery $\frac{1}{2}L' + \frac{1}{2}L''$; but conditional on tails, the two compound lotteries give identical results. The independence axiom requires the sensible conclusion that $\frac{1}{2}L + \frac{1}{2}L''$ be at least as good as $\frac{1}{2}L' + \frac{1}{2}L''$.

The independence axiom is at the heart of the theory of choice under uncertainty. It is unlike anything encountered in the formal theory of preference-based choice discussed in Chapter 1 or its applications in Chapters 3 to 5. This is so precisely because it exploits, in a fundamental manner, the structure of uncertainty present in the model. In the theory of consumer demand, for example, there is no reason to believe that a consumer's preferences over various bundles of goods 1 and 2 should be independent of the quantities of the other goods that he will consume. In the present context, however, it *is* natural to think that a decision maker's preference between two lotteries, say L and L', should determine which of the two he prefers to have as part of a compound lottery *regardless* of the other possible outcome of this compound lottery, say L''. This other outcome L'' should be irrelevant to his choice because, in contrast with the consumer context, he does not consume L or L' together with L'' but, rather, only *instead* of it (if L or L' is the realized outcome).

Exercise 6.B.1: Show that if the preferences $\succsim$ over $\mathscr{L}$ satisfy the independence axiom, then for all $\alpha \in (0, 1)$ and $L, L', L'' \in \mathscr{L}$ we have

$$L \succ L' \quad \text{if and only if} \quad \alpha L + (1-\alpha)L'' \succ \alpha L' + (1-\alpha)L''$$

and

$$L \sim L' \quad \text{if and only if} \quad \alpha L + (1-\alpha)L'' \sim \alpha L' + (1-\alpha)L''.$$

Show also that if $L \succ L'$ and $L'' \succ L'''$, then $\alpha L + (1-\alpha)L'' \succ \alpha L' + (1-\alpha)L'''$.

As we will see shortly, the independence axiom is intimately linked to the representability of preferences over lotteries by a utility function that has an *expected utility form*. Before obtaining that result, we define this property and study some of its features.

Definition 6.B.5: The utility function $U\colon \mathscr{L} \to \mathbb{R}$ has an *expected utility form* if there is an assignment of numbers $(u_1, \ldots, u_N)$ to the N outcomes such that for every simple lottery $L = (p_1, \ldots, p_N) \in \mathscr{L}$ we have

$$U(L) = u_1 p_1 + \cdots + u_N p_N.$$

A utility function $U\colon \mathscr{L} \to \mathbb{R}$ with the expected utility form is called a *von Neumann–Morgenstern (v.N–M) expected utility function.*

Observe that if we let L^n denote the lottery that yields outcome n with probability one, then $U(L^n) = u_n$. Thus, the term *expected utility* is appropriate because with the v.N–M expected utility form, the utility of a lottery can be thought of as the expected value of the utilities u_n of the N outcomes.

The expression $U(L) = \sum_n u_n p_n$ is a general form for a *linear function in the probabilities* $(p_1, \ldots, p_N)$. This linearity property suggests a useful way to think about the expected utility form.

Proposition 6.B.1: A utility function $U\colon \mathscr{L} \to \mathbb{R}$ has an expected utility form if and only if it is *linear*, that is, if and only if it satisfies the property that

$$U\left(\sum_{k=1}^{K} \alpha_k L_k\right) = \sum_{k=1}^{K} \alpha_k U(L_k) \tag{6.B.1}$$

for any K lotteries $L_k \in \mathscr{L}$, $k = 1, \ldots, K$, and probabilities $(\alpha_1, \ldots, \alpha_K) \geq 0$, $\sum_k \alpha_k = 1$.

Proof: Suppose that $U(\cdot)$ satisfies property (6.B.1). We can write any $L = (p_1, \ldots, p_N)$ as a convex combination of the degenerate lotteries $(L^1, \ldots, L^N)$, that is, $L = \sum_n p_n L^n$. We have then $U(L) = U(\sum_n p_n L^n) = \sum_n p_n U(L^n) = \sum_n p_n u_n$. Thus, $U(\cdot)$ has the expected utility form.

In the other direction, suppose that $U(\cdot)$ has the expected utility form, and consider any compound lottery $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$, where $L_k = (p_1^k, \ldots, p_N^k)$. Its reduced lottery is $L' = \sum_k \alpha_k L_k$. Hence,

$$U\left(\sum_k \alpha_k L_k\right) = \sum_n u_n\left(\sum_k \alpha_k p_n^k\right) = \sum_k \alpha_k\left(\sum_n u_n p_n^k\right) = \sum_k \alpha_k U(L_k).$$

Thus, property (6.B.1) is satisfied. ■

The expected utility property is a *cardinal* property of utility functions defined on the space of lotteries. In particular, the result in Proposition 6.B.2 shows that the expected utility form is preserved only by increasing *linear* transformations.

Proposition 6.B.2: Suppose that $U\colon \mathscr{L} \to \mathbb{R}$ is a v.N–M expected utility function for the preference relation $\succsim$ on $\mathscr{L}$. Then $\tilde{U}\colon \mathscr{L} \to \mathbb{R}$ is another v.N–M utility function for $\succsim$ if and only if there are scalars $\beta > 0$ and γ such that $\tilde{U}(L) = \beta U(L) + \gamma$ for every $L \in \mathscr{L}$.

Proof: Begin by choosing two lotteries $\bar{L}$ and $\underline{L}$ with the property that $\bar{L} \succsim L \succsim \underline{L}$ for all $L \in \mathscr{L}$.[6] If $\bar{L} \sim \underline{L}$, then every utility function is a constant and the result follows immediately. Therefore, we assume from now on that $\bar{L} \succ \underline{L}$.

6. These best and worst lotteries can be shown to exist. We could, for example, choose a maximizer and a minimizer of the linear, hence continuous, function $U(\cdot)$ on the simplex of probabilities, a compact set.

Note first that if $U(\cdot)$ is a v.N–M expected utility function and $\tilde{U}(L) = \beta U(L) + \gamma$, then

$$\begin{aligned}\tilde{U}\left(\sum_{k=1}^{K} \alpha_k L_k\right) &= \beta U\left(\sum_{k=1}^{K} \alpha_k L_k\right) + \gamma \\ &= \beta\left[\sum_{k=1}^{K} \alpha_k U(L_k)\right] + \gamma \\ &= \sum_{k=1}^{K} \alpha_k[\beta U(L_k) + \gamma] \\ &= \sum_{k=1}^{K} \alpha_k \tilde{U}(L_k).\end{aligned}$$

Since $\tilde{U}(\cdot)$ satisfies property (6.B.1), it has the expected utility form.

For the reverse direction, we want to show that if both $\tilde{U}(\cdot)$ and $U(\cdot)$ have the expected utility form, then constants $\beta > 0$ and γ exist such that $\tilde{U}(L) = \beta U(L) + \gamma$ for all $L \in \mathscr{L}$. To do so, consider any lottery $L \in \mathscr{L}$, and define $\lambda_L \in [0, 1]$ by

$$U(L) = \lambda_L U(\bar{L}) + (1 - \lambda_L)U(\underline{L}).$$

Thus

$$\lambda_L = \frac{U(L) - U(\underline{L})}{U(\bar{L}) - U(\underline{L})} \tag{6.B.2}$$

Since $\lambda_L U(\bar{L}) + (1 - \lambda_L)U(\underline{L}) = U(\lambda_L \bar{L} + (1 - \lambda_L)\underline{L})$ and $U(\cdot)$ represents the preferences $\succsim$, it must be that $L \sim \lambda_L \bar{L} + (1 - \lambda_L)\underline{L}$. But if so, then since $\tilde{U}(\cdot)$ is also linear and represents these same preferences, we have

$$\begin{aligned}\tilde{U}(L) &= \tilde{U}(\lambda_L \bar{L} + (1 - \lambda_L)\underline{L}) \\ &= \lambda_L \tilde{U}(\bar{L}) + (1 - \lambda_L)\tilde{U}(\underline{L}) \\ &= \lambda_L(\tilde{U}(\bar{L}) - \tilde{U}(\underline{L})) + \tilde{U}(\underline{L}).\end{aligned}$$

Substituting for λ_L from (6.B.2) and rearranging terms yields the conclusion that $\tilde{U}(L) = \beta U(L) + \gamma$, where

$$\beta = \frac{\tilde{U}(\bar{L}) - \tilde{U}(\underline{L})}{U(\bar{L}) - U(\underline{L})}$$

and

$$\gamma = \tilde{U}(\underline{L}) - U(\underline{L})\frac{\tilde{U}(\bar{L}) - \tilde{U}(\underline{L})}{U(\bar{L}) - U(\underline{L})}.$$

This completes the proof ■

A consequence of Proposition 6.B.2 is that for a utility function with the expected utility form, differences of utilities have meaning. For example, if there are four outcomes, the statement "the difference in utility between outcomes 1 and 2 is greater than the difference between outcomes 3 and 4," $u_1 - u_2 > u_3 - u_4$, is equivalent to

$$\tfrac{1}{2}u_1 + \tfrac{1}{2}u_4 > \tfrac{1}{2}u_2 + \tfrac{1}{2}u_3.$$

Therefore, the statement means that the lottery $L = (\frac{1}{2}, 0, 0, \frac{1}{2})$ is preferred to the lottery $L' = (0, \frac{1}{2}, \frac{1}{2}, 0)$. This ranking of utility differences is preserved by all linear transformations of the v.N–M expected utility function.

Note that if a preference relation $\succsim$ on $\mathscr{L}$ is representable by a utility function $U(\cdot)$ that has the expected utility form, then since a linear utility function is continuous, it follows that $\succsim$ is continuous on $\mathscr{L}$. More importantly, the preference relation $\succsim$ must also satisfy the independence axiom. You are asked to show this in Exercise 6.B.2.

Exercise 6.B.2: Show that if the preference relation $\succsim$ on $\mathscr{L}$ is represented by a utility function $U(\cdot)$ that has the expected utility form, then $\succsim$ satisfies the independence axiom.

The expected utility theorem, the central result of this section, tells us that the converse is also true.

The Expected Utility Theorem

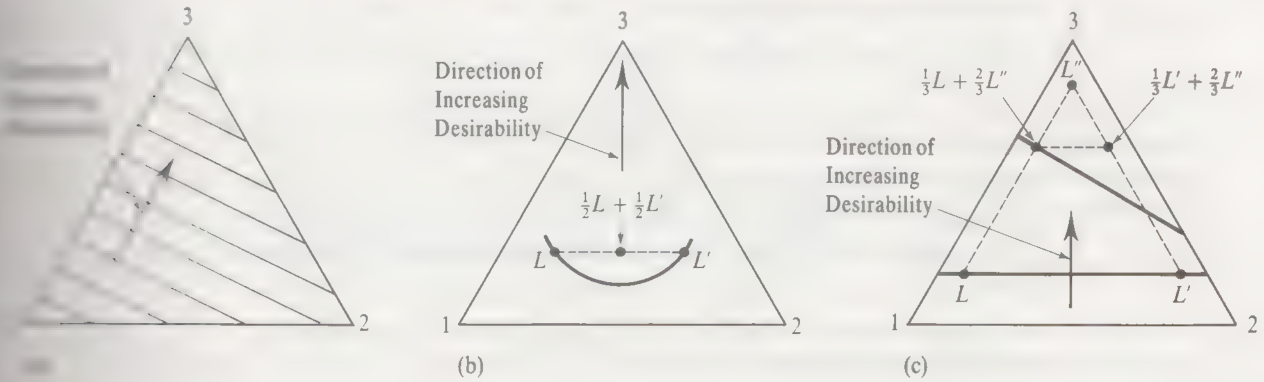
The *expected utility theorem* says that if the decision maker's preferences over lotteries satisfy the continuity and independence axioms, then his preferences are representable by a utility function with the expected utility form. It is the most important result in the theory of choice under uncertainty, and the rest of the book bears witness to its usefulness.

Before stating and proving the result formally, however, it may be helpful to attempt an intuitive understanding of why it is true.

Consider the case where there are only three outcomes. As we have already observed, the continuity axiom insures that preferences on lotteries can be represented by some utility function. Suppose that we represent the indifference map in the simplex, as in Figure 6.B.5. Assume, for simplicity, that we have a conventional map with one-dimensional indifference curves. Because the expected utility form is linear in the probabilities, representability by the expected utility form is equivalent to these indifference curves being straight, parallel lines (you should check this). Figure 6.B.5(a) exhibits an indifference map satisfying these properties. We now argue that these properties are, in fact, consequences of the independence axiom.

Indifference curves are straight lines if, for every pair of lotteries L, L', we have that $L \sim L'$ implies $\alpha L + (1-\alpha)L' \sim L$ for all $\alpha \in [0,1]$. Figure 6.B.5(b) depicts a situation where the indifference curve is not a straight line; we have $L' \sim L$ but

Figure 6.B.5 Geometric explanation of the expected utility theorem. (a) $\succsim$ is representable by a utility function with the expected utility form. (b) Contradiction of the independence axiom. (c) Contradiction of the independence axiom.



$\frac{1}{2}L' + \frac{1}{2}L \succ L$. This is equivalent to saying that

$$\tfrac{1}{2}L' + \tfrac{1}{2}L \succ \tfrac{1}{2}L + \tfrac{1}{2}L. \qquad (6.B.3)$$

But since $L \sim L'$, the independence axiom implies that we must have $\frac{1}{2}L' + \frac{1}{2}L \sim \frac{1}{2}L + \frac{1}{2}L$ (see Exercise 6.B.1). This contradicts (6.B.3), and so we must conclude that indifference curves are straight lines.

Figure 6.B.5(c) depicts two straight but nonparallel indifference lines. A violation of the independence axiom can be constructed in this case, as indicated in the figure. There we have $L \succsim L'$ (in fact, $L \sim L'$), but $\frac{1}{3}L + \frac{2}{3}L'' \succsim \frac{1}{3}L' + \frac{2}{3}L''$ does not hold for the lottery L'' shown in the figure. Thus, indifference curves must be parallel, straight lines if preferences satisfy the independence axiom.

In Proposition 6.B.3, we formally state and prove the expected utility theorem.

Proposition 6.B.3: (*Expected Utility Theorem*) Suppose that the rational preference relation $\succsim$ on the space of lotteries $\mathscr{L}$ satisfies the continuity and independence axioms. Then $\succsim$ admits a utility representation of the expected utility form. That is, we can assign a number u_n to each outcome $n = 1, \ldots, N$ in such a manner that for any two lotteries $L = (p_1, \ldots, p_N)$ and $L' = (p'_1, \ldots, p'_N)$, we have

$$L \succsim L \quad \text{if and only if} \quad \sum_{n=1}^{N} u_n p_n \geq \sum_{n=1}^{N} u_n p'_n. \qquad (6.B.4)$$

Proof: We organize the proof in a succession of steps. For simplicity, we assume that there are best and worst lotteries in $\mathscr{L}$, $\bar{L}$ and $\underline{L}$ (so, $\bar{L} \succsim L \succsim \underline{L}$ for any $L \in \mathscr{L}$).[7] If $\bar{L} \sim \underline{L}$, then all lotteries in $\mathscr{L}$ are indifferent and the conclusion of the proposition holds trivially. Hence, from now on, we assume that $\bar{L} \succ \underline{L}$.

Step 1. If $L \succ L'$ and $\alpha \in (0, 1)$, then $L \succ \alpha L + (1 - \alpha)L' \succ L'$.

This claim makes sense. A nondegenerate mixture of two lotteries will hold a preference position strictly intermediate between the positions of the two lotteries. Formally, the claim follows from the independence axiom. In particular, since $L \succ L'$, the independence axiom implies that (recall Exercise 6.B.1)

$$L = \alpha L + (1 - \alpha)L \succ \alpha L + (1 - \alpha)L' \succ \alpha L' + (1 - \alpha)L' = L'.$$

Step 2. Let $\alpha, \beta \in [0, 1]$. Then $\beta\bar{L} + (1 - \beta)\underline{L} \succ \alpha\bar{L} + (1 - \alpha)\underline{L}$ if and only if $\beta > \alpha$.

Suppose that $\beta > \alpha$. Note first that we can write

$$\beta\bar{L} + (1 - \beta)\underline{L} = \gamma\bar{L} + (1 - \gamma)[\alpha\bar{L} + (1 - \alpha)\underline{L}],$$

where $\gamma = [(\beta - \alpha)/(1 - \alpha)] \in (0, 1]$. By Step 1, we know that $\bar{L} \succ \alpha\bar{L} + (1 - \alpha)\underline{L}$. Applying Step 1 again, this implies that $\gamma\bar{L} + (1 - \gamma)(\alpha\bar{L} + (1 - \alpha)\underline{L}) \succ \alpha\bar{L} + (1 - \alpha)\underline{L}$, and so we conclude that $\beta\bar{L} + (1 - \beta)\underline{L} \succ \alpha\bar{L} + (1 - \alpha)\underline{L}$.

For the converse, suppose that $\beta \leq \alpha$. If $\beta = \alpha$, we must have $\beta\bar{L} + (1 - \beta)\underline{L} \sim \alpha\bar{L} + (1 - \alpha)\underline{L}$. So suppose that $\beta < \alpha$. By the argument proved in the previous

7. In fact, with our assumption of a finite set of outcomes, this can be established as a consequence of the independence axiom (see Exercise 6.B.3).

paragraph (reversing the roles of α and β), we must then have $\alpha\bar{L} + (1-\alpha)\underline{L} \succ \beta\bar{L} + (1-\beta)\underline{L}$.

Step 3. For any $L \in \mathscr{L}$, there is a unique α_L such that $[\alpha_L\bar{L} + (1-\alpha_L)\underline{L}] \sim L$.

Existence of such an α_L is implied by the continuity of $\succsim$ and the fact that $\bar{L}$ and $\underline{L}$ are, respectively, the best and the worst lottery. Uniqueness follows from the result of Step 2.

The existence of α_L is established in a manner similar to that used in the proof of Proposition 3.C.1. Specifically, define the sets

$$\{\alpha \in [0,1]: \alpha\bar{L} + (1-\alpha)\underline{L} \succsim L\} \quad \text{and} \quad \{\alpha \in [0,1]: L \succsim \alpha\bar{L} + (1-\alpha)\underline{L}\}.$$

By the continuity and completeness of $\succsim$, both sets are closed, and any $\alpha \in [0,1]$ belongs to at least one of the two sets. Since both sets are nonempty and $[0,1]$ is connected, it follows that there is some α belonging to both. This establishes the existence of an α_L such that $\alpha_L\bar{L} + (1-\alpha_L)\underline{L} \sim L$.

Step 4. The function $U: \mathscr{L} \to \mathbb{R}$ that assigns $U(L) = \alpha_L$ for all $L \in \mathscr{L}$ represents the preference relation $\succsim$.

Observe that, by Step 3, for any two lotteries $L, L' \in \mathscr{L}$, we have

$$L \succsim L' \quad \text{if and only if} \quad \alpha_L\bar{L} + (1-\alpha_L)\underline{L} \succsim \alpha_{L'}\bar{L} + (1-\alpha_{L'})\underline{L}.$$

Thus, by Step 2, $L \succsim L'$ if and only if $\alpha_L \geq \alpha_{L'}$.

Step 5. The utility function $U(\cdot)$ that assigns $U(L) = \alpha_L$ for all $L \in \mathscr{L}$ is linear and therefore has the expected utility form.

We want to show that for any $L, L' \in \mathscr{L}$, and $\beta \in [0,1]$, we have $U(\beta L + (1-\beta)L') = \beta U(L) + (1-\beta)U(L')$. By definition, we have

$$L \sim U(L)\bar{L} + (1 - U(L))\underline{L}$$

and

$$L' \sim U(L')\bar{L} + (1 - U(L'))\underline{L}.$$

Therefore, by the independence axiom (applied twice),

$$\begin{aligned} \beta L + (1-\beta)L' &\sim \beta[U(L)\bar{L} + (1-U(L))\underline{L}] + (1-\beta)L' \\ &\sim \beta[U(L)\bar{L} + (1-U(L))\underline{L}] + (1-\beta)[U(L')\bar{L} + (1-U(L'))\underline{L}]. \end{aligned}$$

Rearranging terms, we see that the last lottery is algebraically identical to the lottery

$$[\beta U(L) + (1-\beta)U(L')]\bar{L} + [1 - \beta U(L) - (1-\beta)U(L')]\underline{L}.$$

In other words, the compound lottery that gives lottery $[U(L)\bar{L} + (1-U(L))\underline{L}]$ with probability β and lottery $[U(L')\bar{L} + (1-U(L'))\underline{L}]$ with probability $(1-\beta)$ has the same reduced lottery as the compound lottery that gives lottery $\bar{L}$ with probability $[\beta U(L) + (1-\beta)U(L')]$ and lottery $\underline{L}$ with probability $[1 - \beta U(L) - (1-\beta)U(L')]$. Thus

$$\beta L + (1-\beta)L' \sim [\beta U(L) + (1-\beta)U(L')]\bar{L} + [1 - \beta U(L) - (1-\beta)U(L')]\underline{L}.$$

By the construction of $U(\cdot)$ in Step 4, we therefore have

$$U(\beta L + (1-\beta)L') = \beta U(L) + (1-\beta)U(L'),$$

as we wanted.

Together, Steps 1 to 5 establish the existence of a utility function representing $\succsim$ that has the expected utility form. ■

Discussion of the Theory of Expected Utility

A first advantage of the expected utility theorem is technical: It is extremely convenient analytically. This, more than anything else, probably accounts for its pervasive use in economics. It is very easy to work with expected utility and very difficult to do without it. As we have already noted, the rest of the book attests to the importance of the result. Later in this chapter, we will explore some of the analytical uses of expected utility.

A second advantage of the theorem is normative: Expected utility may provide a valuable guide to action. People often find it hard to think systematically about risky alternatives. But if an individual believes that his choices should satisfy the axioms on which the theorem is based (notably, the independence axiom), then the theorem can be used as a guide in his decision process. This point is illustrated in Example 6.B.1.

Example 6.B.1: *Expected Utility as a Guide to Introspection.* A decision maker may not be able to assess his preference ordering between the lotteries L and L' depicted in Figure 6.B.6. The lotteries are too close together, and the differences in the probabilities involved are too small to be understood. Yet, if the decision maker believes that his preferences should satisfy the assumptions of the expected utility theorem, then he may consider L'' instead, which is on the straight line spanned by L and L' but at a significant distance from L. The lottery L'' may not be a feasible choice, but if he determines that $L'' \succ L$, then he can conclude that $L' \succ L$. Indeed, if $L'' \succ L$, then there is an indifference curve separating these two lotteries, as shown in the figure, and it follows from the fact that indifference curves are a family of parallel straight lines that there is also an indifference curve separating L' and L, so that $L' \succ L$. Note that this type of inference is not possible using only the general

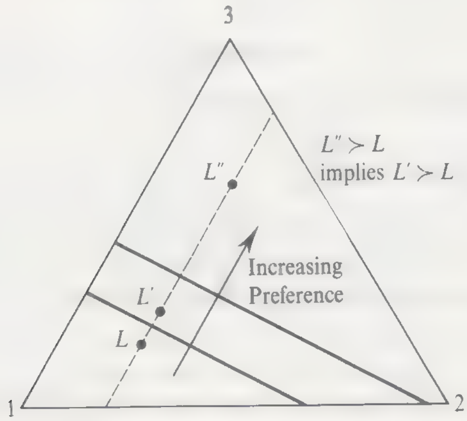


Figure 6.B.6
Expected utilit[y] ... guide to intr[...]

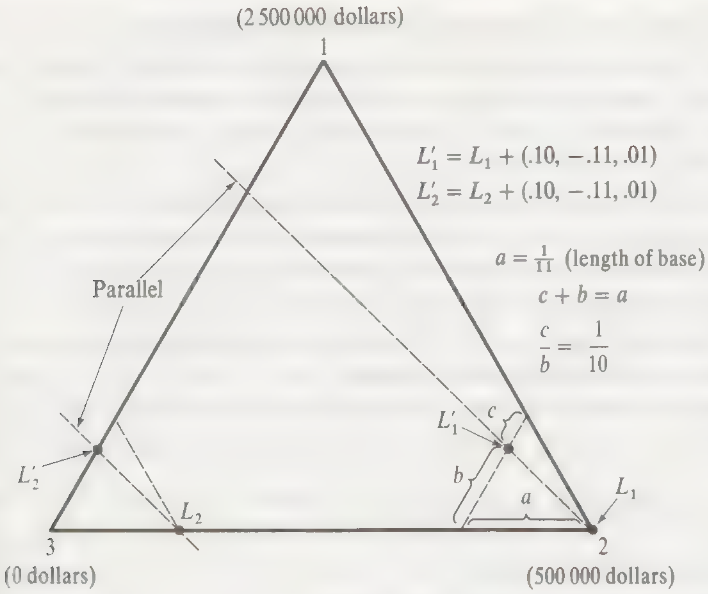


Figure 6.B.7
Depiction of the Allais paradox in the simplex.

utility theory of Chapter 1 because, without the hypotheses of the expected utility theorem, the indifference curves need not be straight lines (with a general indifference map, we could perfectly well have $L'' \succ L$ and $L \succ L'$).

A concrete example of this use of the expected utility theorem is developed in Exercise 6.B.4. ■

As a descriptive theory, however, the expected utility theorem (and, by implication, its central assumption, the independence axiom), is not without difficulties. Examples 6.B.2 and 6.B.3 are designed to test its plausibility.

Example 6.B.2: *The Allais Paradox.* This example, known as the Allais paradox [from Allais (1953)], constitutes the oldest and most famous challenge to the expected utility theorem. It is a thought experiment. There are three possible monetary prizes (so the number of outcomes is $N = 3$):

First Prize	*Second Prize*	*Third Prize*
2 500 000 dollars	500 000 dollars	0 dollars

The decision maker is subjected to two choice tests. The first consists of a choice between the lotteries L_1 and L_1':

$$L_1 = (0, 1, 0) \qquad L_1' = (.10, .89, .01).$$

The second consists of a choice between the lotteries L_2 and L_2':

$$L_2 = (0, .11, .89) \qquad L_2' = (.10, 0, .90).$$

The four lotteries involved are represented in the simplex diagram of Figure 6.B.7.

It is common for individuals to express the preferences $L_1 \succ L_1'$ and $L_2' \succ L_2$.[8]

8. In our classroom experience, roughly half the students choose this way.

The first choice means that one prefers the certainty of receiving 500 000 dollars over a lottery offering a 1/10 probability of getting five times more but bringing with it a tiny risk of getting nothing. The second choice means that, all things considered, a 1/10 probability of getting 2 500 000 dollars is preferred to getting only 500 000 dollars with the slightly better odds of 11/100.

However, these choices are not consistent with expected utility. This can be seen in Figure 6.B.7: The straight lines connecting L_1 to L'_1 and L_2 to L'_2 are parallel. Therefore, if an individual has a linear indifference curve that lies in such a way that L_1 is preferred to L'_1, then a parallel linear indifference curve must make L_2 preferred to L'_2, and vice versa. Hence, choosing L_1 and L'_2 is inconsistent with preferences satisfying the assumptions of the expected utility theorem.

More formally, suppose that there was a v.N–M expected utility function. Denote by u_{25}, u_{05}, and u_0 the utility values of the three outcomes. Then the choice $L_1 \succ L'_1$ implies

$$u_{05} > (.10)u_{25} + (.89)u_{05} + (.01)u_0.$$

Adding $(.89)u_0 - (.89)u_{05}$ to both sides, we get

$$(.11)u_{05} + (.89)u_0 > (.10)u_{25} + (.90)u_0,$$

and therefore any individual with a v.N–M utility function must have $L_2 \succ L'_2$. ■

There are four common reactions to the Allais paradox. The first, propounded by J. Marshack and L. Savage, goes back to the normative interpretation of the theory. It argues that choosing under uncertainty is a reflective activity in which one should be ready to correct mistakes if they are proven inconsistent with the basic principles of choice embodied in the independence axiom (much as one corrects arithmetic mistakes).

The second reaction maintains that the Allais paradox is of limited significance for economics as a whole because it involves payoffs that are out of the ordinary and probabilities close to 0 and 1.

A third reaction seeks to accommodate the paradox with a theory that defines preferences over somewhat larger and more complex objects than simply the ultimate lottery over outcomes. For example, the decision maker may value not only what he receives but also what he receives compared with what he might have received by choosing differently. This leads to *regret theory*. In the example, we could have $L_1 \succ L'_1$ because the expected regret caused by the possibility of getting zero in lottery L'_1, when choosing L_1 would have assured 500 000 dollars, is too great. On the other hand, with the choice between L_2 and L'_2, no such clear-cut regret potential exists; the decision maker was very likely to get nothing anyway.

The fourth reaction is to stick with the original choice domain of lotteries but to give up the independence axiom in favor of something weaker. Exercise 6.B.5 develops this point further.

Example 6.B.3: *Machina's paradox.* Consider the following three outcomes: "a trip to Venice," "watching an excellent movie about Venice," and "staying home." Suppose that you prefer the first to the second and the second to the third.

Now you are given the opportunity to choose between two lotteries. The first lottery gives "a trip to Venice" with probability 99.9% and "watching an excellent movie about Venice" with probability 0.1%. The second lottery gives "a trip to

Venice," again with probability 99.9% and "staying home" with probability 0.1%. The independence axiom forces you to prefer the first lottery to the second. Yet, it would be understandable if you did otherwise. Choosing the second lottery is the rational thing to do if you anticipate that in the event of not getting the trip to Venice, your tastes over the other two outcomes will change: You will be severely *disappointed* and will feel miserable watching a movie about Venice.

The idea of disappointment has parallels with the idea of regret that we discussed in connection with the Allais paradox, but it is not quite the same. Both ideas refer to the influence of "what might have been" on the level of well-being experienced, and it is because of this that they are in conflict with the independence axiom. But disappointment is more directly concerned with what might have been if another outcome of a given lottery had come up, whereas regret should be thought of as regret over a choice not made. ■

Because of the phenomena illustrated in the previous two examples, the search for a useful theory of choice under uncertainty that does not rely on the independence axiom has been an active area of research [see Machina (1987) and also Hey and Orme (1994)]. Nevertheless, the use of the expected utility theorem is pervasive in economics.

An argument sometimes made against the practical significance of violations of the independence axiom is that individuals with such preferences would be weeded out of the marketplace because they would be open to the acceptance of so-called "Dutch books," that is, deals leading to a sure loss of money. Suppose, for example, that there are three lotteries such that $L \succ L'$ and $L \succ L''$ but, in violation of the independence axiom, $\alpha L' + (1-\alpha)L'' \succ L$ for some $\alpha \in (0,1)$. Then, when the decision maker is in the initial position of owning the right to lottery L, he would be willing to pay a small fee to trade L for a compound lottery yielding lottery L' with probability α and lottery L'' with probability $(1-\alpha)$. But as soon as the first stage of this lottery is over, giving him either L' or L'' we could get him to pay a fee to trade this lottery for L. Hence, at that point, he would have paid the two fees but would otherwise be back to his original position.

This may well be a good argument for convexity of the not-better-than sets of $\succsim$, that is, for it to be the case that $L \succsim \alpha L' + (1-\alpha)L''$ whenever $L \succsim L'$ and $L \succsim L''$. This property is implied by the independence axiom but is weaker than it. Dutch book arguments for the full independence axiom are possible, but they are more contrived [see Green (1987)].

Finally, one must use some caution in applying the expected utility theorem because in many practical situations the final outcomes of uncertainty are influenced by actions taken by individuals. Often, these actions should be explicitly modeled but are not. Example 6.B.4 illustrates the difficulty involved.

Example 6.B.4: *Induced preferences.* You are invited to a dinner where you may be served fish (F) or meat (M). You would like to do the proper thing by showing up with white wine if F is served and red wine if M is served. The action of buying the wine must be taken *before* the uncertainty is resolved.

Suppose now that the cost of the bottle of red or white wine is the same and that you are also indifferent between F and M. If you think of the possible outcomes as F and M, then you are apparently indifferent between the lottery that gives F with certainty and the lottery that gives M with certainty. The independence axiom would

then seem to require that you also be indifferent to a lottery that gives F or M with probability $\frac{1}{2}$ each. But you would clearly not be indifferent, since knowing that either F or M will be served with certainty allows you to buy the right wine, whereas, if you are not certain, you will either have to buy both wines or else bring the wrong wine with probability $\frac{1}{2}$.

Yet this example does not contradict the independence axiom. To appeal to the axiom, the decision framework must be set up so that the satisfaction derived from an outcome does not depend on any action taken by the decision maker before the uncertainty is resolved. *Thus, preferences should not be induced or derived from ex ante actions.*[9] Here, the action "acquisition of a bottle of wine" is taken before the uncertainty about the meal is resolved.

To put this situation into the framework required, we must include the ex ante action as part of the description of outcomes. For example, here there would be four outcomes: "bringing red wine when served M," "bringing white wine when served M," "bringing red wine when served F," and "bringing white wine when served F." For any underlying uncertainty about what will be served, you induce a lottery over these outcomes by your choice of action. In this setup, it is quite plausible to be indifferent among "having meat and bringing red wine," "having fish and bringing white wine," or any lottery between these two outcomes, as the independence axiom requires. ■

Although it is not a contradiction to the postulates of expected utility theory, and therefore it is not a serious conceptual difficulty, the induced preferences example nonetheless raises a practical difficulty in the use of the theory. The example illustrates the fact that, in applications, many economic situations do not fit the pure framework of expected utility theory. Preferences are almost always, to some extent, induced.[10]

The expected utility theorem does impose some structure on induced preferences. For example, suppose the complete set of outcomes is $B \times A$, where $B = \{b_1, \dots, b_N\}$ is the set of possible realizations of an exogenous randomness and A is the decision maker's set of possible (ex ante) actions. Under the conditions of the expected utility theorem, for every $a \in A$ and $b_n \in B$, we can assign some utility value $u_n(a)$ to the outcome (b_n, a). Then, for every exogenous lottery $L = (p_1, \dots, p_N)$ on B, we can define a derived utility function by maximizing expected utility:

$$U(L) = \operatorname*{Max}_{a \in A} \sum_n p_n u_n(a).$$

In Exercise 6.B.6, you are asked to show that while $U(L)$, a function on $\mathscr{L}$, need not be linear,

9. Actions taken ex post do not create problems. For example, suppose that $u_n(a_n)$ is the utility derived from outcome n when action a_n is taken after the realization of uncertainty. The decision maker therefore chooses a_n to solve $\text{Max}_{a_n \in A_n} u_n(a_n)$, where A_n is the set of possible actions when outcome n occurs. We can then let $u_n = \text{Max}_{a_n \in A_n} u_n(a_n)$ and evaluate lotteries over the N outcomes as in expected utility theory.

10. Consider, for example, preferences for lotteries over amounts of money available tomorrow. Unless the individual's preferences over consumption today and tomorrow are additively separable, his decision of how much to consume today—a decision that must be made before the resolution of the uncertainty concerning tomorrow's wealth—affects his preferences over these lotteries in a manner that conflicts with the fulfillment of the independence axiom.

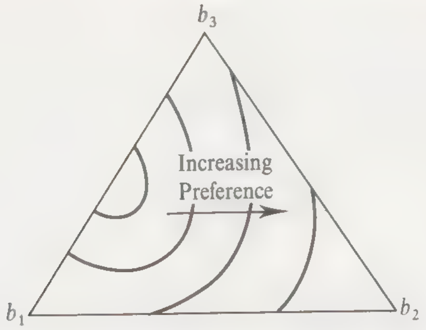


Figure 6.B.8
An indifference map for induced preferences over lotteries on $B = \{b_1, b_2, b_3\}$.

it is nonetheless always *convex*; that is,

$$U(\alpha L + (1 - \alpha)L') \leq \alpha U(L) + (1 - \alpha)U(L').$$

Figure 6.B.8 represents an indifference map for induced preferences in the probability simplex for a case where $N = 3$.

6.C Money Lotteries and Risk Aversion

In many economic settings, individuals seem to display aversion to risk. In this section, we formalize the notion of *risk aversion* and study some of its properties.

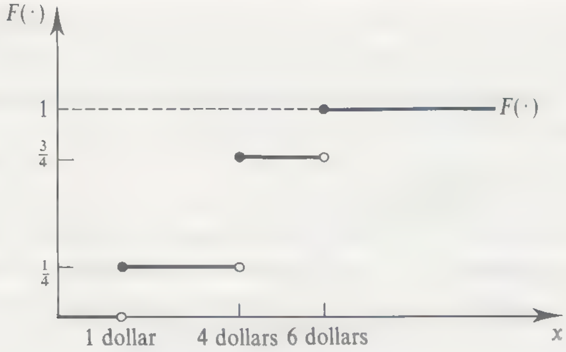
From this section through the end of the chapter, we concentrate on risky alternatives whose outcomes are amounts of money. It is convenient, however, when dealing with monetary outcomes, to treat money as a continuous variable. Strictly speaking, the derivation of the expected utility representation given in Section 6.B assumed a finite number of outcomes. However, the theory can be extended, with some minor technical complications, to the case of an infinite domain. We begin by briefly discussing this extension.

Lotteries over Monetary Outcomes and the Expected Utility Framework

Suppose that we denote amounts of money by the continuous variable x. We can describe a monetary lottery by means of a *cumulative distribution function* $F: \mathbb{R} \to [0, 1]$. That is, for any x, $F(x)$ is the probability that the realized payoff is less than or equal to x. Note that if the distribution function of a lottery has a density function $f(\cdot)$ associated with it, then $F(x) = \int_{-\infty}^{x} f(t)\,dt$ for all x. The advantage of a formalism based on distribution functions over one based on density functions, however, is that the first is completely general. It does not exclude a priori the possibility of a discrete set of outcomes. For example, the distribution function of a lottery with only three monetary outcomes receiving positive probability is illustrated in Figure 6.C.1.

Note that distribution functions preserve the linear structure of lotteries (as do density functions). For example, the final distribution of money, $F(\cdot)$, induced by a compound lottery $(L_1, \ldots, L_K; \alpha_1, \ldots, \alpha_K)$ is just the weighted average of the distributions induced by each of the lotteries that constitute it: $F(x) = \sum_k \alpha_k F_k(x)$, where $F_k(\cdot)$ is the distribution of the payoff under lottery L_k.

From this point on, we shall work with distribution functions to describe lotteries over monetary outcomes. We therefore take the lottery space $\mathscr{L}$ to be the *set of all*



$$\left.\begin{cases} \text{Prob(1 dollar)} = \frac{1}{4} \\ \text{Prob(4 dollars)} = \frac{1}{2} \\ \text{Prob(6 dollars)} = \frac{1}{4} \end{cases}\right\} \rightarrow F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{4} & \text{if } x \in [1, 4) \\ \frac{3}{4} & \text{if } x \in [4, 6) \\ 1 & \text{if } x \geq 6 \end{cases}$$

Figure 6.C.1
A distributio

distribution functions over nonnegative amounts of money, or, more generally, over an interval $[a, +\infty)$.

As in Section 6.B, we begin with a decision maker who has rational preferences $\succsim$ defined over $\mathscr{L}$. The application of the expected utility theorem to outcomes defined by a continuous variable tells us that under the assumptions of the theorem, there is an assignment of utility values $u(x)$ to nonnegative amounts of money with the property that any $F(\cdot)$ can be evaluated by a utility function $U(\cdot)$ of the form

$$U(F) = \int u(x)\, dF(x). \tag{6.C.1}$$

Expression (6.C.1) is the exact extension of the expected utility form to the current setting. The v.N–M utility function $U(\cdot)$ is the mathematical expectation, over the realizations of x, of the values $u(x)$. The latter takes the place of the values $(u_1, \ldots, u_N)$ used in the discrete treatment of Section 6.B.[11] Note that, as before, $U(\cdot)$ is linear in $F(\cdot)$.

The strength of the expected utility representation is that it preserves the very useful expectation form while making the utility of monetary lotteries sensitive not only to the mean but also to the higher moments of the distribution of the monetary payoffs. (See Exercise 6.C.2 for an illuminating quadratic example.)

It is important to distinguish between the utility function $U(\cdot)$, defined on lotteries, and the utility function $u(\cdot)$ defined on sure amounts of money. For this reason, we call $U(\cdot)$ the *von-Neumann–Morgenstern (v.N–M) expected utility function* and $u(\cdot)$ the *Bernoulli utility function.*[12]

11. Given a distribution function $F(x)$, the expected value of a function $\phi(x)$ is given by $\int \phi(x)\, dF(x)$. When $F(\cdot)$ has an associated density function $f(x)$, this expression is exactly equal to $\int \phi(x) f(x)\, dx$. Note also that for notational simplicity, we do not explicitly write the limits of integration when the integral is over the full range of possible realizations of x.

12. The terminology is not standardized. It is common to call $u(\cdot)$ the v.N–M utility function or the expected utility function. We prefer to have a name that is specific to the $u(\cdot)$ function, and so we call it the Bernoulli function for Daniel Bernoulli, who first used an instance of it.

Although the general axioms of Section 6.B yield the expected utility representation, they place no restrictions whatsoever on the Bernoulli utility function $u(\cdot)$. In large part, the analytical power of the expected utility formulation hinges on specifying the Bernoulli utility function $u(\cdot)$ in such a manner that it captures interesting economic attributes of choice behavior. At the simplest level, it makes sense in the current monetary context to postulate that $u(\cdot)$ is *increasing* and *continuous*.[13] We maintain both of these assumptions from now on.

Another restriction, based on a subtler argument, is the *boundedness* (above and below) of $u(\cdot)$. To argue the plausibility of boundedness above (a similar argument applies for boundedness below), we refer to the famous *St. Petersburg–Menger paradox*. Suppose that $u(\cdot)$ is unbounded, so that for every integer m there is an amount of money x_m with $u(x_m) > 2^m$. Consider the following lottery: we toss a coin repeatedly until tails comes up. If this happens in the mth toss, the lottery gives a monetary payoff of x_m. Since the probability of this outcome is $1/2^m$, the expected utility of this lottery is $\sum_{m=1}^{\infty} u(x_m)(1/2^m) \geq \sum_{m=1}^{\infty} (2^m)(1/2^m) = +\infty$. But this means that an individual should be willing to give up all his wealth for the opportunity to play this lottery, a patently absurd conclusion (how much would you pay?).[14]

The rest of this section concentrates on the important property of *risk aversion*, its formulation in terms of the Bernoulli utility function $u(\cdot)$, and its measurement.[15]

Risk Aversion and Its Measurement

The concept of risk aversion provides one of the central analytical techniques of economic analysis, and it is assumed in this book whenever we handle uncertain situations. We begin our discussion of risk aversion with a general definition that does not presume an expected utility formulation.

Definition 6.C.1: A decision maker is a *risk averter* (or exhibits *risk aversion*) if for any lottery $F(\cdot)$, the degenerate lottery that yields the amount $\int x\, dF(x)$ with certainty is at least as good as the lottery $F(\cdot)$ itself. If the decision maker is always [i.e., for any $F(\cdot)$] indifferent between these two lotteries, we say that he is *risk neutral*. Finally, we say that he is *strictly risk averse* if indifference holds only when the two lotteries are the same [i.e., when $F(\cdot)$ is degenerate].

If preferences admit an expected utility representation with Bernoulli utility function $u(x)$, it follows directly from the definition of risk aversion that the decision maker is risk averse if and only if

$$\int u(x)\, dF(x) \leq u\left(\int x\, dF(x)\right) \quad \text{for all } F(\cdot). \tag{6.C.2}$$

Inequality (6.C.2) is called *Jensen's inequality*, and it is the defining property of a concave function (see Section M.C of the Mathematical Appendix). Hence, in the

---

13. In applications, an exception to continuity is sometimes made at $x = 0$ by setting $u(0) = -\infty$.

14. In practice, most utility functions commonly used are not bounded. Paradoxes are avoided because the class of distributions allowed by the modeler in each particular application is a limited one. Note also that if we insisted on $u(\cdot)$ being defined on $(-\infty, \infty)$ then any nonconstant $u(\cdot)$ could not be both concave and bounded (above and below).

15. Arrow (1971) and Pratt (1964) are the classical references in this area.

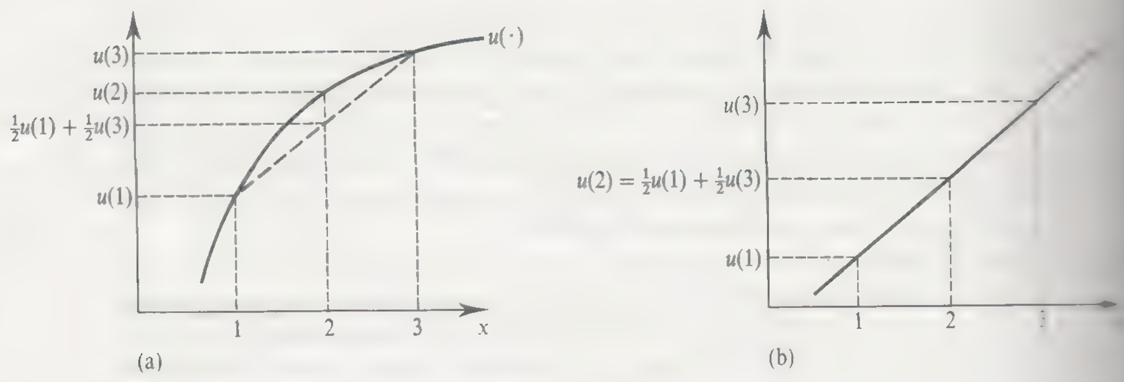


Figure 6.C.2 Risk aversion (a) and risk neutrality (b).

context of expected utility theory, we see that *risk aversion is equivalent to the concavity of* $u(\cdot)$ and that strict risk aversion is equivalent to the strict concavity of $u(\cdot)$. This makes sense. Strict concavity means that the marginal utility of money is decreasing. Hence, at any level of wealth x, the utility gain from an extra dollar is smaller than (the absolute value of) the utility loss of having a dollar less. It follows that a risk of gaining or losing a dollar with even probability is not worth taking. This is illustrated in Figure 6.C.2(a); in the figure we consider a gamble involving the gain or loss of 1 dollar from an initial position of 2 dollars. The (v.N–M)utility of this gamble, $\frac{1}{2}u(1) + \frac{1}{2}u(3)$, is strictly less than that of the initial certain position $u(2)$.

For a risk-neutral expected utility maximizer, (6.C.2) must hold with *equality* for all $F(\cdot)$. Hence, the decision maker is risk neutral if and only if the Bernoulli utility function of money $u(\cdot)$ is linear. Figure 6.C.2(b) depicts the (v.N–M) utility associated with the previous gamble for a risk neutral individual. Here the individual is indifferent between the gambles that yield a mean wealth level of 2 dollars and a certain wealth of 2 dollars. Definition 6.C.2 introduces two useful concepts for the analysis of risk aversion.

Definition 6.C.2: Given a Bernoulli utility function $u(\cdot)$ we define the following concepts:

(i) The *certainty equivalent of* $F(\cdot)$, denoted $c(F, u)$, is the amount of money for which the individual is indifferent between the gamble $F(\cdot)$ and the certain amount $c(F, u)$; that is,

$$u(c(F, u)) = \int u(x)\, dF(x). \tag{6.C.3}$$

(ii) For any fixed amount of money x and positive number ε, the *probability premium* denoted by $\pi(x, \varepsilon, u)$, is the excess in winning probability over fair odds that makes the individual indifferent between the certain outcome x and a gamble between the two outcomes $x + \varepsilon$ and $x - \varepsilon$. That is

$$u(x) = (\tfrac{1}{2} + \pi(x, \varepsilon, u))\, u(x + \varepsilon) + (\tfrac{1}{2} - \pi(x, \varepsilon, u))\, u(x - \varepsilon). \tag{6.C.4}$$

These two concepts are illustrated in Figure 6.C.3. In Figure 6.C.3(a), we exhibit the geometric construction of $c(F, u)$ for an even probability gamble between 1 and 3 dollars. Note that $c(F, u) < 2$, implying that some expected return is traded for certainty. The satisfaction of the inequality $c(F, u) \le \int x\, dF(x)$ for all $F(\cdot)$ is, in fact,

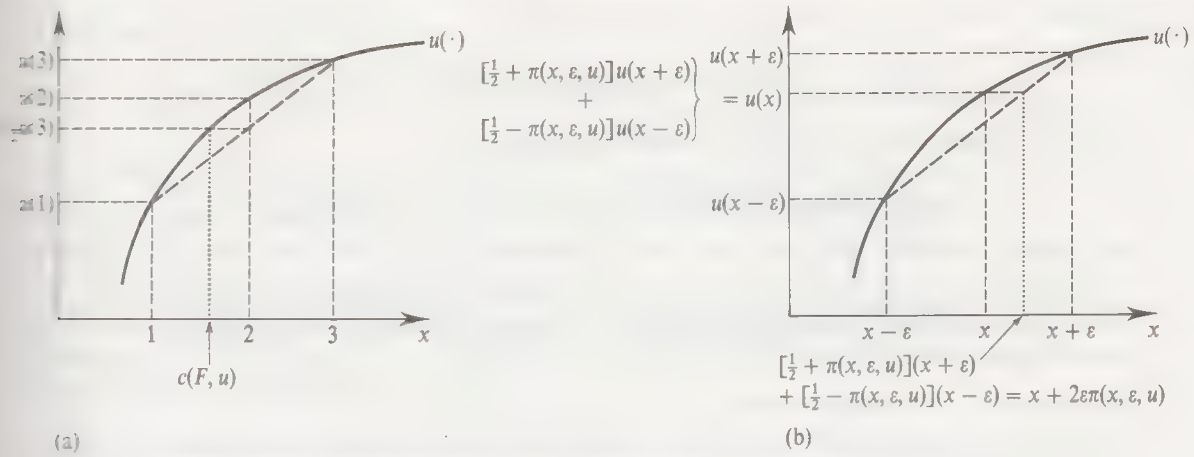


Figure 6.C.3 The certainty equivalent (a) and the probability premium (b).

equivalent to the decision maker being a risk averter. To see this, observe that since $u(\cdot)$ is nondecreasing, we have

$$c(F, u) \leq \int x\, dF(x) \Leftrightarrow u(c(F, u)) \leq u\left(\int x\, dF(x)\right) \Leftrightarrow \int u(x)\, dF(x) \leq u\left(\int x\, dF(x)\right),$$

where the last $\Leftrightarrow$ follows from the definition of $c(F, u)$.

In Figure 6.C.3(b), we exhibit the geometric construction of $\pi(x, \varepsilon, u)$. We see that $\pi(x, \varepsilon, u) > 0$; that is, better than fair odds must be given for the individual to accept the risk. In fact, the satisfaction of the inequality $\pi(x, \varepsilon, u) \geq 0$ for all x and $\varepsilon > 0$ is also equivalent to risk aversion (see Exercise 6.C.3).

These points are formally summarized in Proposition 6.C.1.

Proposition 6.C.1: Suppose a decision maker is an expected utility maximizer with a Bernoulli utility function $u(\cdot)$ on amounts of money. Then the following properties are equivalent:

(i) The decision maker is risk averse.
(ii) $u(\cdot)$ is concave.[16]
(iii) $c(F, u) \leq \int x\, dF(x)$ for all $F(\cdot)$.
(iv) $\pi(x, \varepsilon, u) \geq 0$ for all x, ε.

Examples 6.C.1 to 6.C.3 illustrate the use of the risk aversion concept.

Example 6.C.1: *Insurance.* Consider a strictly risk-averse decision maker who has an initial wealth of w but who runs a risk of a loss of D dollars. The probability of the loss is π. It is possible, however, for the decision maker to buy insurance. One unit of insurance costs q dollars and pays 1 dollar if the loss occurs. Thus, if α units of insurance are bought, the wealth of the individual will be $w - \alpha q$ if there is no loss and $w - \alpha q - D + \alpha$ if the loss occurs. Note, for purposes of later discussion, that the decision maker's expected wealth is then $w - \pi D + \alpha(\pi - q)$. The decision maker's problem is to choose the optimal level of α. His utility maximization problem is

16. Recall that if $u(\cdot)$ is twice differentiable then concavity is equivalent to $u''(x) \leq 0$ for all x.

therefore

$$\underset{\alpha \geq 0}{\text{Max}} \ (1-\pi)u(w-\alpha q) + \pi u(w - \alpha q - D + \alpha).$$

If α^* is an optimum, it must satisfy the first-order condition:

$$-q(1-\pi)u'(w-\alpha^* q) + \pi(1-q)u'(w - D + \alpha^*(1-q)) \leq 0,$$

with equality if $\alpha^* > 0$.

Suppose now that the price q of one unit of insurance is *actuarially fair* in the sense of it being equal to the expected cost of insurance. That is, $q = \pi$. Then the first-order condition requires that

$$u'(w - D + \alpha^*(1-\pi)) - u'(w - \alpha^* \pi) \leq 0,$$

with equality if $\alpha^* > 0$.

Since $u'(w - D) > u'(w)$, we must have $\alpha^* > 0$, and therefore

$$u'(w - D + \alpha^*(1-\pi)) = u'(w - \alpha^* \pi).$$

Because $u'(\cdot)$ is strictly decreasing, this implies

$$w - D + \alpha^*(1-\pi) = w - \alpha^* \pi,$$

or, equivalently,

$$\alpha^* = D.$$

Thus, *if insurance is actuarially fair, the decision maker insures completely*. The individual's final wealth is then $w - \pi D$, regardless of the occurrence of the loss.

This proof of the complete insurance result uses first-order conditions, which is instructive but not really necessary. Note that if $q = \pi$, then the decision maker's expected wealth is $w - \pi D$ for any α. Since setting $\alpha = D$ allows him to reach $w - \pi D$ with certainty, the definition of risk aversion directly implies that this is the optimal level of α. ■

Example 6.C.2: *Demand for a Risky Asset.* An *asset* is a divisible claim to a financial return in the future. Suppose that there are two assets, a safe asset with a return of 1 dollar per dollar invested and a risky asset with a random return of z dollars per dollar invested. The random return z has a distribution function $F(z)$ that we assume satisfies $\int z\, dF(z) > 1$; that is, its mean return exceeds that of the safe asset.

An individual has initial wealth w to invest, which can be divided in any way between the two assets. Let α and β denote the amounts of wealth invested in the risky and the safe asset, respectively. Thus, for any realization z of the random return, the individual's *portfolio* (α, β) pays $\alpha z + \beta$. Of course, we must also have $\alpha + \beta = w$.

The question is how to choose α and β. The answer will depend on $F(\cdot)$, w, and the Bernoulli utility function $u(\cdot)$. The utility maximization problem of the individual is

$$\underset{\alpha, \beta \geq 0}{\text{Max}} \int u(\alpha z + \beta)\, dF(z)$$

$$\text{s.t. } \alpha + \beta = w.$$

Equivalently, we want to maximize $\int u(w + \alpha(z-1))\, dF(z)$ subject to $0 \leq \alpha \leq w$. If

α^* is optimal, it must satisfy the Kuhn–Tucker first-order conditions:[17]

$$\phi(\alpha^*) = \int u'(w + \alpha^*[z - 1])(z - 1)\, dF(z) \begin{cases} \leq 0 & \text{if } \alpha^* < w, \\ \geq 0 & \text{if } \alpha^* > 0. \end{cases}$$

Note that $\int z\, dF(z) > 1$ implies $\phi(0) > 0$. Hence, $\alpha^* = 0$ cannot satisfy this first-order condition. We conclude that the optimal portfolio has $\alpha^* > 0$. The general principle illustrated in this example, is that *if a risk is actuarially favorable, then a risk averter will always accept at least a small amount of it.*

This same principle emerges in Example 6.C.1 if insurance is not actuarially fair. In Exercise 6.C.1, you are asked to show that if $q > \pi$, then the decision maker will not fully insure (i.e., will accept some risk). ■

Example 6.C.3: *General Asset Problem.* In the previous example, we could define the utility $U(\alpha, \beta)$ of the portfolio (α, β) as $U(\alpha, \beta) = \int u(\alpha z + \beta)\, dF(z)$. Note that $U(\cdot)$ is then an increasing, continuous, and concave utility function. We now discuss an important generalization. We assume that we have N assets (one of which may be the safe asset) with asset n giving a return of z_n per unit of money invested. These returns are jointly distributed according to a distribution function $F(z_1, \ldots, z_N)$. The utility of holding a *portfolio* of assets $(\alpha_1, \ldots, \alpha_N)$ is then

$$U(\alpha_1, \ldots, \alpha_N) = \int u(\alpha_1 z_1 + \cdots + \alpha_N z_N)\, dF(z_1, \ldots, z_N).$$

This utility function for portfolios, defined on $\mathbb{R}^N_+$, is also increasing, continuous, and concave (see Exercise 6.C.4). This means that, formally, we can treat assets as the usual type of commodities and apply to them the demand theory developed in Chapters 2 and 3. Observe, in particular, how risk aversion leads to a convex indifference map for portfolios. ■

Suppose that the lotteries pay in vectors of physical goods rather than in money. Formally, the space of outcomes is then the consumption set $\mathbb{R}^L_+$ (all the previous discussion can be viewed as the special case in which there is a single good). In this more general setting, the concept of risk aversion given by Definition 6.C.1 is perfectly well defined. Furthermore, if there is a Bernoulli utility function $u: \mathbb{R}^L_+ \to \mathbb{R}$, then risk aversion is still equivalent to the concavity of $u(\cdot)$. Hence, we have here another justification for the convexity assumption of Chapter 3. Under the assumptions of the expected utility theorem, the convexity of preferences for perfectly certain amounts of the physical commodities must hold if for any lottery with commodity payoffs the individual always prefers the certainty of the mean commodity bundle to the lottery itself.

In Exercise 6.C.5, you are asked to show that if preferences over lotteries with commodity payoffs exhibit risk aversion, then, at given commodity prices, the induced preferences on money lotteries (where consumption decisions are made after the realization of wealth) are also risk averse. Thus, in principle, it is possible to build the theory of risk aversion on the more primitive notion of lotteries over the final consumption of goods.

---

17. The objective function is concave in α because the concavity of $u(\cdot)$ implies that $\int u''(w + \alpha(z - 1))(z - 1)^2\, dF(x) \leq 0$.

The Measurement of Risk Aversion

Now that we know what it means to be risk averse, we can try to measure the extent of risk aversion. We begin by defining one particularly useful measure and discussing some of its properties.

Definition 6.C.3: Given a (twice-differentiable) Bernoulli utility function $u(\cdot)$ for money, the *Arrow Pratt coefficient of absolute risk aversion* at x is defined as $r_A(x) = -u''(x)/u'(x)$.

The Arrow–Pratt measure can be motivated as follows: We know that risk neutrality is equivalent to the linearity of $u(\cdot)$, that is, to $u''(x) = 0$ for all x. Therefore, it seems logical that the degree of risk aversion be related to the *curvature* of $u(\cdot)$. In Figure 6.C.4, for example, we represent two Bernoulli utility functions $u_1(\cdot)$ and $u_2(\cdot)$ normalized (by choice of origin and units) to have the same utility and marginal utility values at wealth level x. The certainty equivalent for a small risk with mean x is smaller for $u_2(\cdot)$ than for $u_1(\cdot)$, suggesting that risk aversion increases with the curvature of the Bernoulli utility function at x. One possible measure of curvature of the Bernoulli utility function $u(\cdot)$ at x is $u''(x)$. However, this is not an adequate measure because it is not invariant to positive linear transformations of the utility function. To make it invariant, the simplest modification is to use $u''(x)/u'(x)$. If we change sign so as to have a positive number for an increasing and concave $u(\cdot)$, we get the Arrow–Pratt measure.

A more precise motivation for $r_A(x)$ as a measure of the degree of risk aversion can be obtained by considering a fixed wealth x and studying the behavior of the probability premium $\pi(x, \varepsilon, u)$ as $\varepsilon \to 0$ [for simplicity, we write it as $\pi(\varepsilon)$]. Differentiating the identity (6.C.4) that defines $\pi(\cdot)$ twice with respect to ε (assume that $\pi(\cdot)$ is differentiable), and evaluating at $\varepsilon = 0$, we get $4\pi'(0)u'(x) + u''(x) = 0$. Hence

$$r_A(x) = 4\pi'(0).$$

Thus, $r_A(x)$ measures the rate at which the probability premium increases at certainty with the small risk measured by ε.[18] As we go along, we will find additional related interpretations of the Arrow–Pratt measure.

18. For a similar derivation relating $r_A(\cdot)$ to the rate of change of the certainty equivalent with respect to a small increase in a small risk around certainty, see Exercise 6.C.20.

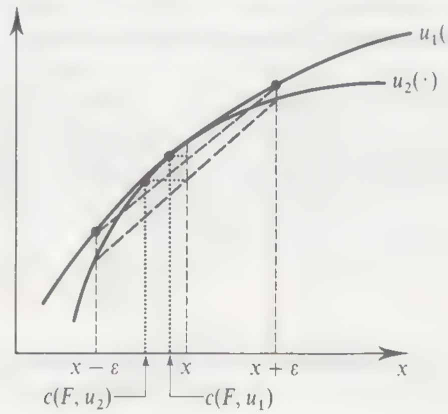


Figure 6.C.4
Differing ...
risk aversion

Note that, up to two integration constants, the utility function $u(\cdot)$ can be recovered from $r_A(\cdot)$ by integrating twice. The integration constants are irrelevant because the Bernoulli utility is identified only up to two constants (origin and units). Thus, the Arrow–Pratt risk aversion measure $r_A(\cdot)$ fully characterizes behavior under uncertainty.

Example 6.C.4: Consider the utility function $u(x) = -e^{-ax}$ for $a > 0$. Then $u'(x) = ae^{-ax}$ and $u''(x) = -a^2e^{-ax}$. Therefore, $r_A(x, u) = a$ for all x. It follows from the observation just made that the general form of a Bernoulli utility function with an Arrow–Pratt measure of absolute risk aversion equal to the constant $a > 0$ at all x is $u(x) = -\alpha e^{-ax} + \beta$ for some $\alpha > 0$ and β. ■

Once we are equipped with a measure of risk aversion, we can put it to use in comparative statics exercises. Two common situations are the comparisons of risk attitudes across individuals with different utility functions and the comparison of risk attitudes for one individual at different levels of wealth.

Comparisons across individuals

Given two Bernoulli utility functions $u_1(\cdot)$ and $u_2(\cdot)$, when can we say that $u_2(\cdot)$ is unambiguously *more risk averse than* $u_1(\cdot)$? Several possible approaches to a definition seem plausible:

(i) $r_A(x, u_2) \geq r_A(x, u_1)$ for every x.

(ii) There exists an increasing concave function $\psi(\cdot)$ such that $u_2(x) = \psi(u_1(x))$ at all x; that is, $u_2(\cdot)$ is a concave transformation of $u_1(\cdot)$. [In other words, $u_2(\cdot)$ is "more concave" than $u_1(\cdot)$.]

(iii) $c(F, u_2) \leq c(F, u_1)$ for any $F(\cdot)$.

(iv) $\pi(x, \varepsilon, u_2) \geq \pi(x, \varepsilon, u_1)$ for any x and ε.

(v) Whenever $u_2(\cdot)$ finds a lottery $F(\cdot)$ at least as good as a riskless outcome $\bar{x}$, then $u_1(\cdot)$ also finds $F(\cdot)$ at least as good as $\bar{x}$. That is, $\int u_2(x)\,dF(x) \geq u_2(\bar{x})$ implies $\int u_1(x)\,dF(x) \geq u_1(\bar{x})$ for any $F(\cdot)$ and $\bar{x}$.[19]

In fact, these five definitions are equivalent.

Proposition 6.C.2: Definitions (i) to (v) of the *more-risk-averse-than* relation are equivalent.

Proof: We will not give a complete proof. (You are asked to establish some of the implications in Exercises 6.C.6 and 6.C.7.) Here we will show the equivalence of (i) and (ii) under differentiability assumptions.

Note, first that we always have $u_2(x) = \psi(u_1(x))$ for some increasing function $\psi(\cdot)$; this is true simply because $u_1(\cdot)$ and $u_2(\cdot)$ are ordinally identical (more money is preferred to less). Differentiating, we get

$$u_2'(x) = \psi'(u_1(x))u_1'(x)$$

and

$$u_2''(x) = \psi'(u_1(x))u_1''(x) + \psi''(u_1(x))(u_1'(x))^2.$$

Dividing both sides of the second expression by $u_2'(x) > 0$, and using the first

19. In other words, any risk that $u_2(\cdot)$ would accept starting from a position of certainty would also be accepted by $u_1(\cdot)$.

expression, we get

$$r_A(x, u_2) = r_A(x, u_1) - \frac{\psi''(u_1(x))}{\psi'(u_1(x))} u_1'(x).$$

Thus, $r_A(x, u_2) \geq r_A(x, u_1)$ for all x if and only if $\psi''(u_1) \leq 0$ for all u_1 in the range of $u_1(\cdot)$. ■

The more-risk-averse-than relation is a *partial ordering* of Bernoulli utility functions; it is transitive but far from complete. Typically, two Bernoulli utility functions $u_1(\cdot)$ and $u_2(\cdot)$ will not be comparable; that is, we will have $r_A(x, u_1) > r_A(x, u_2)$ at some x but $r_A(x', u_1) < r_A(x', u_2)$ at some other $x' \neq x$.

Example 6.C.2 continued: We take up again the asset portfolio problem between a safe and a risky asset discussed in Example 6.C.2. Suppose that we now have two individuals with Bernoulli utility functions $u_1(\cdot)$ and $u_2(\cdot)$, and denote by α_1^* and α_2^* their respective optimal investments in the risky asset. We will show that *if* $u_2(\cdot)$ *is more risk averse than* $u_1(\cdot)$, *then* $\alpha_2^* < \alpha_1^*$; that is, the second decision maker invests less in the risky asset than the first.

To repeat from our earlier discussion, the asset allocation problem for $u_1(\cdot)$ is

$$\underset{0 \leq \alpha \leq w}{\text{Max}} \int u_1(w - \alpha + \alpha z)\, dF(z).$$

Assuming an interior solution, the first-order condition is

$$\int (z - 1) u_1'(w + \alpha_1^*[z - 1])\, dF(z) = 0. \tag{6.C.5}$$

The analogous expression for the utility function $u_2(\cdot)$ is

$$\phi_2(\alpha_2^*) = \int (z - 1) u_2'(w + \alpha_2^*[z - 1])\, dF(z) = 0. \tag{6.C.6}$$

As we know, the concavity of $u_2(\cdot)$ implies that $\phi_2(\cdot)$ is decreasing. Therefore, if we show that $\phi_2(\alpha_1^*) < 0$, it must follow that $\alpha_2^* < \alpha_1^*$, which is the result we want. Now, $u_2(x) = \psi(u_1(x))$ allows us to write

$$\phi_2(\alpha_1^*) = \int (z - 1)\psi'(u_1(w + \alpha_1^*[z - 1]))u_1'(w + \alpha_1^*[z - 1])\, dF(z) < 0. \tag{6.C.7}$$

To understand the final inequality, note that the integrand of expression (6.C.7) is the same as that in (6.C.5) except that it is multiplied by $\psi'(\cdot)$, a positive decreasing function of z [recall that $u_2(\cdot)$ more risk averse than $u_1(\cdot)$ means that the increasing function $\psi(\cdot)$ is concave; that is, $\psi'(\cdot)$ is positive and decreasing]. Hence, the integral (6.C.7) underweights the positive values of $(z - 1)u_1'(w + \alpha_1^*[z - 1])$, which obtain for $z > 1$, relative to the negative values, which obtain for $z < 1$. Since, in (6.C.5), the integral of the positive and the negative parts of the integrand added to zero, they now must add to a negative number. This establishes the desired inequality. ■

Comparisons across wealth levels

It is a common contention that wealthier people are willing to bear more risk than poorer people. Although this might be due to differences in utility functions across people, it is more likely that the source of the difference lies in the possibility that

wealthier people "can afford to take a chance." Hence, we shall explore the implications of the condition stated in Definition 6.C.4.

Definition 6.C.4: The Bernoulli utility function $u(\cdot)$ for money exhibits *decreasing absolute risk aversion* if $r_A(x, u)$ is a decreasing function of x.

Individuals whose preferences satisfy the decreasing absolute risk aversion property take more risk as they become wealthier. Consider two levels of initial wealth $x_1 > x_2$. Denote the increments or decrements to wealth by z. Then the individual evaluates risk at x_1 and x_2 by, respectively, the induced Bernoulli utility functions $u_1(z) = u(x_1 + z)$ and $u_2(z) = u(x_2 + z)$. Comparing an individual's attitudes toward risk as his level of wealth changes is like comparing the utility functions $u_1(\cdot)$ and $u_2(\cdot)$, a problem we have just studied. If $u(\cdot)$ displays decreasing absolute risk aversion, then $r_A(z, u_2) \geq r_A(z, u_1)$ for all z. This is condition (i) of Proposition 6.C.2. Hence, the result in Proposition 6.C.3 follows directly from Proposition 6.C.2.

Proposition 6.C.3: The following properties are equivalent:

(i) The Bernoulli utility function $(\cdot)$ exhibits decreasing absolute risk aversion.

(ii) Whenever $x_2 < x_1$, $u_2(z) = u(x_2 + z)$ is a concave transformation of $u_1(z) = u(x_1 + z)$.

(iii) For any risk $F(z)$, the certainty equivalent of the lottery formed by adding risk z to wealth level x, given by the amount c_x at which $u(c_x) = \int u(x + z)\, dF(z)$, is such that $(x - c_x)$ is decreasing in x. That is, the higher x is, the less is the individual willing to pay to get rid of the risk.

(iv) The probability premium $\pi(x, \varepsilon, u)$ is decreasing in x.

(v) For any $F(z)$, if $\int u(x_2 + z)\, dF(z) \geq u(x_2)$ and $x_2 < x_1$, then $\int u(x_1 + z)\, dF(z) \geq u(x_1)$.

Exercise 6.C.8: Assume that the Bernoulli utility function $u(\cdot)$ exhibits decreasing absolute risk aversion. Show that for the asset demand model of Example 6.C.2 (and Exercise 6.C.2 continued), the optimal allocation between the safe and the risky assets places an increasing amount of wealth in the risky asset as w rises (i.e., the risky asset is a normal good).

The assumption of decreasing absolute risk aversion yields many other economically reasonable results concerning risk-bearing behavior. However, in applications, it is often too weak and, because of its analytical convenience, it is sometimes supplemented by a stronger assumption: *nonincreasing relative risk aversion.*

To understand the concept of relative risk aversion, note that the concept of absolute risk aversion is suited to the comparison of attitudes toward risky projects whose outcomes are *absolute gains or losses* from current wealth. But it is also of interest to evaluate risky projects whose outcomes are *percentage* gains or losses of current wealth. The concept of relative risk aversion does just this.

Let $t > 0$ stand for *proportional* increments or decrements of wealth. Then, an individual with Bernoulli utility function $u(\cdot)$ and initial wealth x can evaluate a random percentage risk by means of the utility function $\tilde{u}(t) = u(tx)$. The initial wealth position corresponds to $t = 1$. We already know that for a small risk around $t = 1$, the degree of risk aversion is well captured by $\tilde{u}''(1)/\tilde{u}'(1)$. Noting that $\tilde{u}''(1)/\tilde{u}'(1) = xu''(x)/u'(x)$, we are led to the concept stated in Definition 6.C.5.

Definition 6.C.5: Given a Bernoulli utility function $u(\cdot)$, the *coefficient of relative risk aversion at x* is $r_R(x, u) = -xu''(x)/u'(x)$.

Consider now how this measure varies with wealth. The property of *nonincreasing relative risk aversion* says that the individual becomes less risk averse with regard to gambles that are proportional to his wealth as his wealth increases. This is a stronger assumption than decreasing absolute risk aversion: Since $r_R(x, u) = xr_A(x, u)$, a risk-averse individual with decreasing relative risk aversion will exhibit decreasing absolute risk aversion, but the converse is not necessarily the case.

As before, we can examine various implications of this concept. Proposition 6.C.4 is an abbreviated parallel to Proposition 6.C.3.

Proposition 6.C.4: The following conditions for a Bernoulli utility function $u(\cdot)$ on amounts of money are equivalent:

(i) $r_R(x, u)$ is decreasing in x.

(ii) Whenever $x_2 < x_1$, $\tilde{u}_2(t) = u(tx_2)$ is a concave transformation of $\tilde{u}_1(t) = u(tx_1)$.

(iii) Given any risk $F(t)$ on $t > 0$, the certainty equivalent $\bar{c}_x$ defined by $u(\bar{c}_x) = \int u(tx)\, dF(t)$ is such that $x/\bar{c}_x$ is decreasing in x.

Proof: Here we show only that (i) implies (iii). To this effect, fix a distribution $F(t)$ on $t > 0$, and, for any x, define $u_x(t) = u(tx)$. Let $c(x)$ be the usual certainty equivalent (from Definition 6.C.2): $u_x(c(x)) = \int u_x(t)\, dF(t)$. Note that $-u_x''(t)/u_x'(t) = -(1/t)tx[u''(tx)/u'(tx)]$ for any x. Hence if (i) holds, then $u_{x'}(\cdot)$ is less risk averse than $u_x(\cdot)$ whenever $x' > x$. Therefore, by Proposition 6.C.2, $c(x') > c(x)$ and we conclude that $c(\cdot)$ is increasing. Now, by the definition of $u_x(\cdot)$, $u_x(c(x)) = u(xc(x))$. Also

$$u_x(c(x)) = \int u_x(t)\, dF(t) = \int u(tx)\, dF(t) = u(\bar{c}_x).$$

Hence, $\bar{c}_x/x = c(x)$, and so $x/\bar{c}_x$ is decreasing. This concludes the proof. ■

Example 6.C.2 continued: In Exercise 6.C.11, you are asked to show that if $r_R(x, u)$ is decreasing in x, then the proportion of wealth invested in the risky asset $\gamma = \alpha/w$ is increasing with the individual's wealth level w. The opposite conclusion holds if $r_R(x, u)$ is increasing in x. If $r_R(x, u)$ is a constant independent of x, then the fraction of wealth invested in the risky asset is independent of w [see Exercise 6.C.12 for the specific analytical form that $u(\cdot)$ must have]. Models with constant relative risk aversion are encountered often in finance theory, where they lead to considerable analytical simplicity. Under this assumption, no matter how the wealth of the economy and its distribution across individuals evolves over time, the portfolio decisions of individuals in terms of budget shares do not vary (as long as the safe return and the distribution of random returns remain unchanged). ■

6.D Comparison of Payoff Distributions in Terms of Return and Risk

In this section, we continue our study of lotteries with monetary payoffs. In contrast with Section 6.C, where we compared utility functions, our aim here is to compare

payoff distributions. There are two natural ways that random outcomes can be compared: according to the level of returns and according to the dispersion of returns. We will therefore attempt to give meaning to two ideas: that of a distribution $F(\cdot)$ yielding unambiguously higher returns than $G(\cdot)$ and that of $F(\cdot)$ being unambiguously less risky than $G(\cdot)$. These ideas are known, respectively, by the technical terms of *first-order stochastic dominance* and *second-order stochastic dominance*.[20]

In all subsequent developments, we restrict ourselves to distributions $F(\cdot)$ such that $F(0) = 0$ and $F(x) = 1$ for some x.

First-Order Stochastic Dominance

We want to attach meaning to the expression: "The distribution $F(\cdot)$ yields unambiguously higher returns than the distribution $G(\cdot)$." At least two sensible criteria suggest themselves. First, we could test whether every expected utility maximizer who values more over less prefers $F(\cdot)$ to $G(\cdot)$. Alternatively, we could verify whether, for every amount of money x, the probability of getting at least x is higher under $F(\cdot)$ than under $G(\cdot)$. Fortunately, these two criteria lead to the same concept.

Definition 6.D.1: The distribution $F(\cdot)$ *first-order stochastically dominates* $G(\cdot)$ if, for every nondecreasing function $u\colon \mathbb{R} \to \mathbb{R}$, we have

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x).$$

Proposition 6.D.1: The distribution of monetary payoffs $F(\cdot)$ first-order stochastically dominates the distribution $G(\cdot)$ if and only if $F(x) \leq G(x)$ for every x.

Proof: Given $F(\cdot)$ and $G(\cdot)$ denote $H(x) = F(x) - G(x)$. Suppose that $H(\bar{x}) > 0$ for some $\bar{x}$. Then we can define a nondecreasing function $u(\cdot)$ by $u(x) = 1$ for $x > \bar{x}$ and $u(x) = 0$ for $x \leq \bar{x}$. This function has the property that $\int u(x)\, dH(x) = -H(\bar{x}) < 0$, and so the "only if" part of the proposition follows.

For the "if" part of the proposition we first put on record, without proof, that it suffices to establish the equivalence for differentiable utility functions $u(\cdot)$. Given $F(\cdot)$ and $G(\cdot)$, denote $H(x) = F(x) - G(x)$. Integrating by parts, we have

$$\int u(x)\, dH(x) = [u(x)H(x)]_0^{\infty} - \int u'(x)H(x)\, dx.$$

Since $H(0) = 0$ and $H(x) = 0$ for large x, the first term of this expression is zero. It follows that $\int u(x)\, dH(x) \geq 0$ [or, equivalently, $\int u(x)\, dF(x) - \int u(x)\, dG(x) \geq 0$] if and only if $\int u'(x)H(x)\, dx \leq 0$. Thus, if $H(x) \leq 0$ for all x and $u(\cdot)$ is increasing, then $\int u'(x)H(x)\, dx \leq 0$ and the "if" part of the proposition follows. ■

In Exercise 6.D.1 you are asked to verify Proposition 6.D.1 for the case of lotteries with three possible outcomes. In Figure 6.D.1, we represent two distributions $F(\cdot)$ and $G(\cdot)$. Distribution $F(\cdot)$ first-order stochastically dominates $G(\cdot)$ because the graph of $F(\cdot)$ is uniformly below the graph of $G(\cdot)$. Note two important points: First, first-order stochastic dominance does *not* imply that every possible return of the

20. They were introduced into economics in Rothschild and Stiglitz (1970).

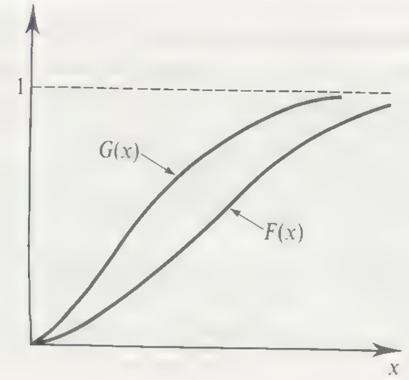


Figure 6.D.
F(·) fir
stoch
domin

superior distribution is larger than every possible return of the inferior one. In the figure, the set of possible outcomes is the same for the two distributions. Second, although $F(\cdot)$ first-order stochastically dominating $G(\cdot)$ implies that the mean of x under $F(\cdot)$, $\int x\,dF(x)$, is greater than its mean under $G(\cdot)$, a ranking of the means of two distributions does *not* imply that one first-order stochastically dominates the other; rather, the entire distribution matters (see Exercise 6.D.3).

Example 6.D.1: Consider a compound lottery that has as its first stage a realization of x distributed according to $G(\cdot)$ and in its second stage applies to the outcome x of the first stage an "upward probabilistic shift." That is, if outcome x is realized in the first stage, then the second stage pays a final amount of money $x + z$, where z is distributed according to a distribution $H_x(z)$ with $H_x(0) = 0$. Thus, $H_x(\cdot)$ generates a *final* return of at least x with probability one. (Note that the distributions applied to different x's may differ.)

Denote the resulting reduced distribution by $F(\cdot)$. Then for any nondecreasing function $u\colon \mathbb{R} \to \mathbb{R}$, we have

$$\int u(x)\,dF(x) = \int\left[\int u(x+z)\,dH_x(z)\right]dG(x) \geq \int u(x)\,dG(x).$$

So $F(\cdot)$ first-order stochastically dominates $G(\cdot)$.

A specific example is illustrated in Figure 6.D.2. As Figure 6.D.2(a) shows, $G(\cdot)$ is an even randomization between 1 and 4 dollars. The outcome "1 dollar" is then

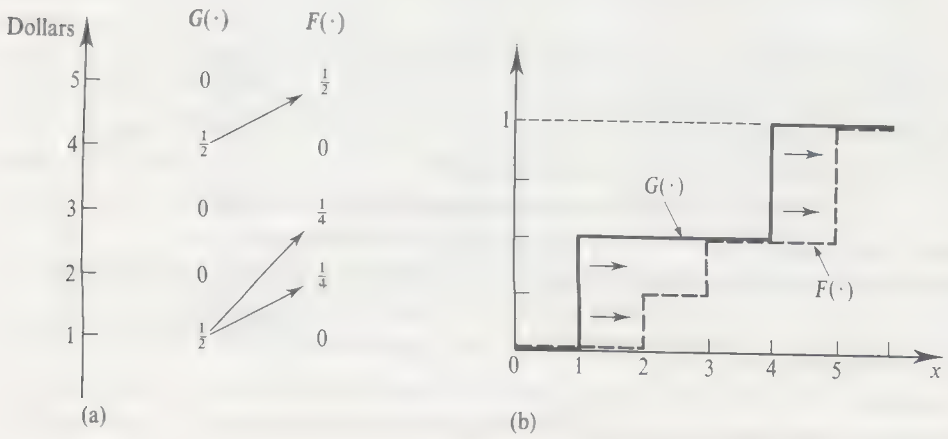


Figure 6.D.2
F(·) fir
stocha
domina

shifted up to an even probability between 2 and 3 dollars, and the outcome "4 dollars" is shifted up to 5 dollars with probability one. Figure 6.D.2(b) shows that $F(x) \leq G(x)$ at all x.

It can be shown that the reverse direction also holds. Whenever $F(\cdot)$ first-order stochastically dominates $G(\cdot)$, it is possible to generate $F(\cdot)$ from $G(\cdot)$ in the manner suggested in this example. Thus, this provides yet another approach to the characterization of the first-order stochastic dominance relation. ■

Second-Order Stochastic Dominance

First-order stochastic dominance involves the idea of "higher/better" vs. "lower/worse." We want next to introduce a comparison based on relative *riskiness* or *dispersion*. To avoid confusing this issue with the trade-off between returns and risk, *we will restrict ourselves for the rest of this section to comparing distributions with the same mean.*

Once again, a definition suggests itself: Given two distributions $F(\cdot)$ and $G(\cdot)$ with the same mean [that is, with $\int x\, dF(x) = \int x\, dG(x)$], we say that $G(\cdot)$ is riskier than $F(\cdot)$ if every risk averter prefers $F(\cdot)$ and $G(\cdot)$. This is stated formally in Definition 6.D.2.

Definition 6.D.2: For any two distributions $F(x)$ and $G(\cdot)$ with the same mean, $F(\cdot)$ *second-order stochastically dominates* (or *is less risky than*) $G(\cdot)$ if for every nondecreasing concave function $u\colon \mathbb{R}_+ \to \mathbb{R}$, we have

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x).$$

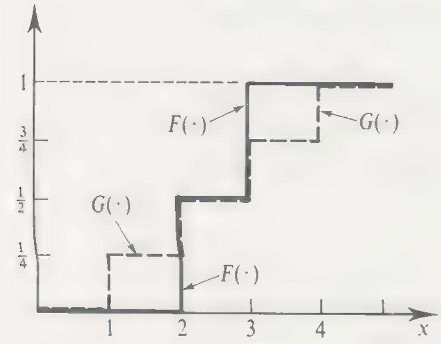
Example 6.D.2 introduces an alternative way to characterize the second-order stochastic dominance relation.

Example 6.D.2: *Mean-Preserving Spreads.* Consider the following compound lottery: In the first stage, we have a lottery over x distributed according to $F(\cdot)$. In the second stage, we randomize each possible outcome x further so that the final payoff is $x + z$, where z has a distribution function $H_x(z)$ with a mean of zero [i.e., $\int z\, dH_x(z) = 0$]. Thus, the mean of $x + z$ is x. Let the resulting reduced lottery be denoted by $G(\cdot)$. When lottery $G(\cdot)$ can be obtained from lottery $F(\cdot)$ in this manner for some distribution $H_x(\cdot)$, we say that $G(\cdot)$ is a *mean-preserving spread* of $F(\cdot)$.

For example, $F(\cdot)$ may be an even probability distribution between 2 and 3 dollars. In the second step we may spread the 2 dollars outcome to an even probability between 1 and 3 dollars, and the 3 dollars outcome to an even probability between 2 and 4 dollars. Then $G(\cdot)$ is the distribution that assigns probability $\frac{1}{4}$ to the four outcomes: 1, 2, 3, 4 dollars. These two distributions $F(\cdot)$ and $G(\cdot)$ are depicted in Figure 6.D.3.

The type of two-stage operation just described keeps the mean of $G(\cdot)$ equal to that of $F(\cdot)$. In addition, if $u(\cdot)$ is concave, we can conclude that

$$\int u(x)\, dG(x) = \int \left(\int u(x+z)\, dH_x(z) \right) dF(x) \leq \int u\left(\int (x+z)\, dH_x(z) \right) dF(x)$$
$$= \int u(x)\, dF(x),$$



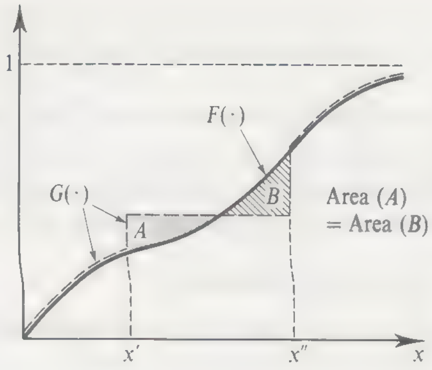


Figure 6.D.3 (left) $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$.

Figure 6.D.4 (right) $G(\cdot)$ is an [illegible] increase in risk [illegible] $F(\cdot)$.

and so $F(\cdot)$ second-order stochastically dominates $G(\cdot)$. It turns out that the converse is also true: If $F(\cdot)$ second-order stochastically dominates $G(\cdot)$, then $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$. Hence, *saying that $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$ is equivalent to saying that $F(\cdot)$ second-order stochastically dominates $G(\cdot)$.* ■

Example 6.D.3 provides another illustration of a mean-preserving spread.

Example 6.D.3: *An Elementary Increase in Risk.* We say that $G(\cdot)$ constitutes an *elementary increase in risk* from $F(\cdot)$ if $G(\cdot)$ is generated from $F(\cdot)$ by taking all the mass that $F(\cdot)$ assigns to an interval $[x', x'']$ and transferring it to the endpoints x' and x'' in such a manner that the mean is preserved. This is illustrated in Figure 6.D.4. An elementary increase in risk is a mean-preserving spread. [In Exercise 6.D.3, you are asked to verify directly that if $G(\cdot)$ is an elementary increase in risk from $F(\cdot)$, then $F(\cdot)$ second-order stochastically dominates $G(\cdot)$.] ■

We can develop still another way to capture the second-order stochastic dominance idea. Suppose that we have two distributions $F(\cdot)$ and $G(\cdot)$ with the same mean. Recall that, for simplicity, we assume that $F(\bar{x}) = G(\bar{x}) = 1$ for some $\bar{x}$. Integrating by parts (and recalling the equality of the means) yields

$$\int_0^{\bar{x}} (F(x) - G(x))\, dx = -\int_0^{\bar{x}} x\, d(F(x) - G(x)) + (F(\bar{x}) - G(\bar{x}))\bar{x} = 0. \qquad (6.D.1)$$

That is, the areas below the two distribution functions are the same over the interval $[0, \bar{x}]$. Because of this fact, the regions marked A and B in Figure 6.D.4 must have the same area. Note that for the two distributions in the figure, this implies that

$$\int_0^x G(t)\, dt \geq \int_0^x F(t)\, dt \quad \text{for all } x. \qquad (6.D.2)$$

It turns out that property (6.D.2) is equivalent to $F(\cdot)$ second-order stochastically dominating $G(\cdot)$.[21] As an application, suppose that $F(\cdot)$ and $G(\cdot)$ have the same mean and that the graph of $G(\cdot)$ is initially above the graph of $F(\cdot)$ and then moves

21. We will not prove this. The claim can be established along the same lines used to prove Proposition 6.D.1 except that we must integrate by parts twice and take into account expression (6.D.1).

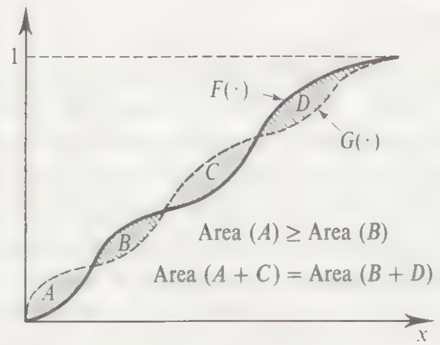


Figure 6.D.5
$F(\cdot)$ second-order stochastically dominates $G(\cdot)$.

permanently below it (as in Figures 6.D.3 and 6.D.4). Then because of (6.D.1), condition (6.D.2) must be satisfied, and we can conclude that $G(\cdot)$ is riskier than $F(\cdot)$. As a more elaborate example, consider Figure 6.D.5, which shows two distributions having the same mean and satisfying (6.D.2). To verify that (6.D.2) is satisfied, note that area A has been drawn to be at least as large as area B and that the equality of the means [i.e., (6.D.1)] implies that the areas $B + D$ and $A + C$ must be equal.

We state Proposition 6.D.2 without proof.

Proposition 6.D.2: Consider two distributions $F(\cdot)$ and $G(\cdot)$ with the same mean. Then the following statements are equivalent:

(i) $F(\cdot)$ second-order stochastically dominates $G(\cdot)$.
(ii) $G(\cdot)$ is a mean-preserving spread of $F(\cdot)$.
(iii) Property (6.D.2) holds.

In Exercise 6.D.4, you are asked to verify the equivalence of these three properties in the probability simplex diagram.

6.E State-dependent Utility

In this section, we consider an extension of the analysis presented in the preceding two sections. In Sections 6.C and 6.D, we assumed that the decision maker cares solely about the distribution of monetary payoffs he receives. This says, in essence, that the underlying cause of the payoff is of no importance. If the cause is one's state of health, however, this assumption is unlikely to be fulfilled.[22] The distribution function of monetary payoffs is then not the appropriate object of individual choice. Here we consider the possibility that the decision maker may care not only about his monetary returns but also about the underlying events, or *states of nature*, that cause them.

We begin by discussing a convenient framework for modeling uncertain alternatives that, in contrast to the lottery apparatus, recognizes underlying states of nature. (We will encounter it repeatedly throughout the book, especially in Chapter 19.)

22. On the other hand, if it is an event such as the price of some security in a portfolio, the assumption is more likely to be a good representation of reality.

State-of-Nature Representations of Uncertainty

In Sections 6.C and 6.D, we modeled a risky alternative by means of a distribution function over monetary outcomes. Often, however, we know that the random outcome is generated by some underlying causes. A more detailed description of uncertain alternatives is then possible. For example, the monetary payoff of an insurance policy might depend on whether or not a certain accident has happened, the payoff on a corporate stock on whether the economy is in a recession, and the payoff of a casino gamble on the number selected by the roulette wheel.

We call these underlying causes *states*, or *states of nature*. We denote the set of states by S and an individual state by $s \in S$. For simplicity, we assume here that the set of states is finite and that each state s has a well-defined, objective probability $\pi_s > 0$ that it occurs. We abuse notation slightly by also denoting the total number of states by S.

An uncertain alternative with (nonnegative) monetary returns can then be described as a function that maps realizations of the underlying state of nature into the set of possible money payoffs $\mathbb{R}_+$. Formally, such a function is known as a *random variable*.

Definition 6.E.1: A *random variable* is a function $g: S \to \mathbb{R}_+$ that maps states into monetary outcomes.[23]

Every random variable $g(\cdot)$ gives rise to a money lottery describable by the distribution function $F(\cdot)$ with $F(x) = \sum_{\{s: g(s) \leq x\}} \pi_s$ for all x. Note that there is a loss in information in going from the random variable representation of uncertainty to the lottery representation; we do not keep track of which states give rise to a given monetary outcome, and only the aggregate probability of every monetary outcome is retained.

Because we take S to be finite, we can represent a random variable with monetary payoffs by the vector $(x_1, \ldots, x_S)$, where x_s is the nonnegative monetary payoff in state s. The set of all nonnegative random variables is then $\mathbb{R}^S_+$.

State-Dependent Preferences and the Extended Expected Utility Representation

The primitive datum of our theory is now a rational preference relation $\succsim$ on the set $\mathbb{R}^S_+$ of nonnegative random variables. Note that this formal setting is parallel to the one developed in Chapters 2 to 4 for consumer choice. The similarity is not merely superficial. If we define commodity s as the random variable that pays one dollar if and only if state s occurs (this is called a *contingent commodity* in Chapter 19), then the set of nonnegative random variables $\mathbb{R}^S_+$ is precisely the set of nonnegative bundles of these S contingent commodities.

As we shall see, it is very convenient if, in the spirit of the previous sections of this chapter, we can represent the individual's preferences over monetary outcomes by a utility function that possesses an *extended expected utility form.*

---

23. For concreteness, we restrict the outcomes to be nonnegative amounts of money. As we did in Section 6.B, we could equally well use an abstract outcome set C instead of $\mathbb{R}_+$.

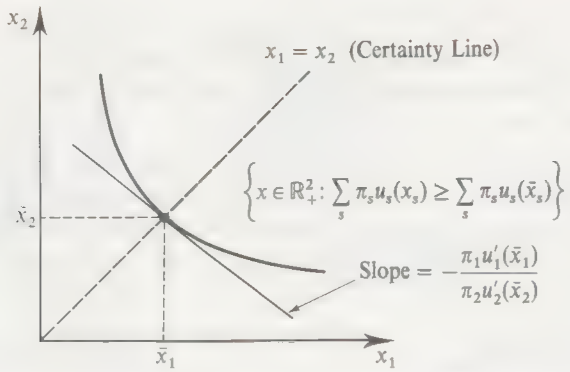


Figure 6.E.1
State-dependent preferences.

Definition 6.E.2: The preference relation $\succsim$ has an *extended expected utility representation* if for every $s \in S$, there is a function $u_s: \mathbb{R}_+ \to \mathbb{R}$ such that for any $(x_1, \ldots, x_S) \in \mathbb{R}^S_+$ and $(x_1', \ldots, x_S') \in \mathbb{R}^S_+$,

$$(x_1, \ldots, x_S) \succsim (x_1', \ldots, x_S') \quad \text{if and only if} \quad \sum_s \pi_s u_s(x_s) \geq \sum_s \pi_s u_s(x_s').$$

To understand Definition 6.E.2, recall the analysis in Section 6.B. If only the distribution of money payoffs mattered, and if preferences on money distributions satisfied the expected utility axioms, then the expected utility theorem leads to a *state-independent* (we will also say *state-uniform*) expected utility representation $\sum_s \pi_s u(x_s)$, where $u(\cdot)$ is the Bernoulli utility function on amounts of money.[24] The generalization in Definition 6.E.2 allows for a different function $u_s(\cdot)$ in every state.

Before discussing the conditions under which an extended utility representation exists, we comment on its usefulness as a tool in the analysis of choice under uncertainty. This usefulness is primarily a result of the behavior of the indifference sets around the *money certainty line*, the set of random variables that pay the same amount in every state. Figure 6.E.1 depicts state-dependent preferences in the space $\mathbb{R}^2_+$ for a case where $S = 2$ and the $u_s(\cdot)$ functions are concave (as we shall see later, concavity of these functions follows from risk aversion considerations). The certainty line in Figure 6.E.1 is the set of points with $x_1 = x_2$. The marginal rate of substitution at a point $(\bar{x}, \bar{x})$ is $\pi_1 u_1'(\bar{x})/\pi_2 u_2'(\bar{x})$. Thus, the slope of the indifference curves on the certainty line reflects the nature of state dependence as well as the probabilities of the different states. In contrast, with state-uniform (i.e., identical across states) utility functions, the marginal rate of substitution at any point on the certainty line equals the ratio of the probabilities of the states (implying that this slope is the same at all points on the certainty line).

Example 6.E.1: *Insurance with State-dependent Utility.* One interesting implication of state dependency arises when actuarially fair insurance is available. Suppose there are two states: State 1 is the state where no loss occurs, and state 2 is the state where a loss occurs. (This economic situation parallels that in Example 6.C.1.) The individual's initial situation (i.e., in the absence of any insurance purchase) is a

24. Note that the random variable $(x_1, \ldots, x_S)$ induces a money lottery that pays x_s with probability π_s. Hence, $\sum_s \pi_s u(x_s)$ is its expected utility.

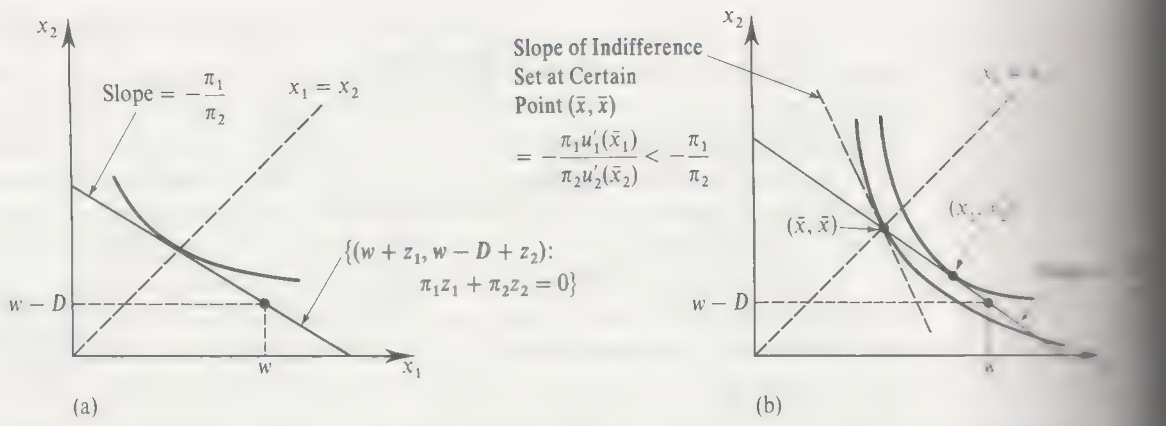


Figure 6.E.2 Insurance purchase with state-dependent utility. (a) State-uniform utility. (b) State-de...

random variable $(w, w - D)$ that gives the individual's wealth in the two states. This is depicted in Figure 6.E.2(a). We can represent an insurance contract by a random variable $(z_1, z_2) \in \mathbb{R}^2$ specifying the net change in wealth in the two states (the insurance payoff in the state less any premiums paid). Thus, if the individual purchases insurance contract (z_1, z_2), his final wealth position will be $(w + z_1, w - D + z_2)$. The insurance policy (z_1, z_2) is actuarially fair if its expected payoff is zero, that is, if $\pi_1 z_1 + \pi_2 z_2 = 0$.

Figure 6.E.2(a) shows the optimal insurance purchase when a risk-averse expected utility maximizer with state-uniform preferences can purchase any actuarially fair insurance policy he desires. His budget set is the straight line drawn in the figure. We saw in Example 6.C.2 that under these conditions, a decision maker with state-uniform utility would insure completely. This is confirmed here because if there is no state dependency, the budget line is tangent to an indifference curve at the certainty line.

Figure 6.E.2(b) depicts the situation with state-dependent preferences. The decision maker will now prefer a point such as (x_1', x_2') to the certain outcome $(\bar{x}, \bar{x})$. This creates a desire to have a higher payoff in state 1, where $u_1'(\cdot)$ is relatively higher, in exchange for a lower payoff in state 2. ■

Existence of an Extended Expected Utility Representation

We now investigate conditions for the existence of an extended expected utility representation.

Observe first that since $\pi_s > 0$ for every s, we can formally include π_s in the definition of the utility function at state s. That is, to find an extended expected utility representation, it suffices that there be functions $u_s(\cdot)$ such that

$$(x_1, \ldots, x_S) \succsim (x_1', \ldots, x_S') \quad \text{if and only if} \quad \sum_s u_s(x_s) \geq \sum_s u_s(x_s').$$

This is because if such functions $u_s(\cdot)$ exist, then we can define $\tilde{u}_s(\cdot) = (1/\pi_s)u_s(\cdot)$ for each $s \in S$, and we will have $\sum_s u_s(x_s) \geq \sum_s u_s(x_s')$ if and only if $\sum_s \pi_s \tilde{u}_s(x_s) \geq \sum_s \pi_s \tilde{u}_s(x_s')$. Thus, from now on, we focus on the existence of an additively separable form $\sum_s u_s(\cdot)$, and the π_s's cease to play any role in the analysis.

It turns out that the extended expected utility representation can be derived *in exactly the same way* as the expected utility representation of Section 6.B if we appropriately enlarge the domain over which preferences are defined.[25] Accordingly, we now allow for the possibility that within each state s, the monetary payoff is not a certain amount of money x_s but a random amount with distribution function $F_s(\cdot)$. We denote these uncertain alternatives by $L = (F_1, \ldots, F_S)$. Thus, L is a kind of compound lottery that assigns well-defined monetary gambles as prizes contingent on the realization of the state of the world s. We denote by $\mathscr{L}$ the set of all such possible lotteries. in 15, 16, 19

Our starting point is now a rational preference relation $\succsim$ on $\mathscr{L}$. Note that $\alpha L + (1-\alpha)L' = (\alpha F_1 + (1-\alpha)F'_1, \ldots, \alpha F_S + (1-\alpha)F'_S)$ has the usual interpretation as the reduced lottery arising from a randomization between L and L', although here we are dealing with a reduced lottery within each state s. Hence, we can appeal to the same logic as in Section 6.B and impose an independence axiom on preferences.

Definition 6.E.3: The preference relation $\succsim$ on $\mathscr{L}$ satisfies the *extended independence axiom* if for all $L, L', L'' \in \mathscr{L}$ and $\alpha \in (0, 1)$, we have

$$L \succsim L' \quad \text{if and only if} \quad \alpha L + (1-\alpha)L'' \succsim \alpha L' + (1-\alpha)L''.$$

We also make a continuity assumption: Except for the reinterpretation of $\mathscr{L}$, this *continuity axiom* is exactly the same as that in Section 6.B; we refer to Definition 6.B.3 for its statement.

Proposition 6.E.1: (*Extended Expected Utility Theorem*) Suppose that the preference relation $\succsim$ on the space of lotteries $\mathscr{L}$ satisfies the continuity and extended independence axioms. Then we can assign a utility function $u_s(\cdot)$ for money in every state s such that for any $L = (F_1, \ldots, F_S)$ and $L' = (F'_1, \ldots, F'_S)$, we have

$$L \succsim L' \quad \text{if and only if} \quad \sum_s \left(\int u_s(x_s)\, dF_s(x_s) \right) \geq \sum_s \left(\int u_s(x_s)\, dF'_s(x_s) \right).$$

Proof: The proof is identical, almost word for word, to the proof of the expected utility theorem (Proposition 6.B.2).

Suppose, for simplicity, that we restrict ourselves to a finite number $\{x_1, \ldots, x_N\}$ of monetary outcomes. Then we can identify the set $\mathscr{L}$ with Δ^S, where Δ is the $(N-1)$-dimensional simplex. Our aim is to show that $\succsim$ can be represented by a linear utility function $U(L)$ on Δ^S. To see this, note that, up to an additive constant that can be neglected, $U(p_1^1, \ldots, p_N^1, \ldots, p_1^S, \ldots, p_N^S)$ is a linear function of its arguments if it can be written as $U(L) = \sum_{n,s} u_{n,s} p_n^s$ for some values $u_{n,s}$. In this case, we can write $U(L) = \sum_s (\sum_n u_{n,s} p_n^s)$, which, letting $u_s(x_n) = u_{n,s}$, is precisely the form of a utility function on $\mathscr{L}$ that we want.

Choose $\bar{L}$ and $\underline{L}$ such that $\bar{L} \succsim L \succsim \underline{L}$ for all $L \in \mathscr{L}$. As in the proof of Proposition 6.B.2, we can then define $U(L)$ by the condition

$$L \sim U(L)\bar{L} + (1 - U(L))\underline{L}.$$

Applying the extended independence axiom in exactly the same way as we applied the independence axiom in the proof of Proposition 6.B.2 yields the result that $U(L)$ is indeed a linear utility function on $\mathscr{L}$. ■

---

25. By pushing the enlargement further than we do here, it would even be possible to view the existence of an extended utility representation as a corollary of the expected utility theorem.

Proposition 6.F.1 gives us a utility representation $\sum_s u_s(x_s)$ for the preferences on state-by-state sure outcomes $(x_1, \ldots, x_S) \in \mathbb{R}^S_+$ that has two properties. First, it is additively separable across states. Second, every $u_s(\cdot)$ is a Bernoulli utility function that can be used to evaluate lotteries over money payoffs in state s by means of expected utility. It is because of the second property that risk aversion (defined in exactly the same manner as in Section 6.C) is equivalent to the concavity of each $u_s(\cdot)$.

There is another approach to the extended expected utility representation that rests with the preferences $\succsim$ defined on $\mathbb{R}^S_+$ and does not appeal to preferences defined on a larger space. It is based on the so-called *sure-thing axiom.*

Definition 6.E.4: The preference relation $\succsim$ satisfies the *sure-thing axiom* if, for any subset of states $E \subset S$ (E is called an *event*), whenever $(x_1, \ldots, x_S)$ and $(x'_1, \ldots, x'_S)$ differ only in the entries corresponding to E (so that $x'_s = x_s$ for $s \notin E$), the preference ordering between $(x_1, \ldots, x_S)$ and $(x'_1, \ldots, x'_S)$ is independent of the particular (common) payoffs for states not in E. Formally, suppose that $(x_1, \ldots, x_S)$, $(x'_1, \ldots, x'_S)$, $(\bar{x}_1, \ldots, \bar{x}_S)$, and $(\bar{x}'_1, \ldots, \bar{x}'_S)$ are such that

$$\text{For all } s \notin E\text{:} \quad x_s = x'_s \quad \text{and} \quad \bar{x}_s = \bar{x}'_s.$$
$$\text{For all } s \in E\text{:} \quad x_s = \bar{x}_s \quad \text{and} \quad x'_s = \bar{x}'_s.$$

Then $(\bar{x}_1, \ldots, \bar{x}_S) \succsim (\bar{x}'_1, \ldots, \bar{x}'_S)$ if and only if $(x_1, \ldots, x_S) \succsim (x'_1, \ldots, x'_S)$.

The intuitive content of this axiom is similar to that of the independence axiom. It simply says that if two random variables cannot be distinguished in the complement of E, then the ordering among them can depend only on the values they take on E. In other words, tastes conditional on an event should not depend on what the payoffs would have been in states that have not occurred.

If $\succsim$ admits an extended expected utility representation, the sure-thing axiom holds because then $(x_1, \ldots, x_S) \succsim (x'_1, \ldots, x'_S)$ if and only if $\sum_s (u_s(x_s) - u_s(x'_s)) \geq 0$, and any term of the sum with $x_s = x'_s$ will cancel. In the other direction we have Proposition 6.E.2.

Proposition 6.E.2: Suppose that there are at least three states and that the preferences $\succsim$ on $\mathbb{R}^S_+$ are continuous and satisfy the sure-thing axiom. Then $\succsim$ admits an extended expected utility representation.

Idea of Proof: A complete proof is too advanced to be given in any detail. One wants to show that under the assumptions, preferences admit an additively separable utility representation $\sum_s u_s(x_s)$. This is not easy to show, and it is not a result particularly related to uncertainty. The conditions for the existence of an additively separable utility function for continuous preferences on the positive orthant of a Euclidean space (i.e., the context of Chapter 3) are well understood; as it turns out, they are *formally identical* to the sure-thing axiom (see Exercise 3.G.4). ■

Although the sure-thing axiom yields an extended expected utility representation $\sum_s \pi_s u_s(x_s)$, we would emphasize that randomizations over monetary payoffs in a state s have not been considered in this approach, and therefore we cannot bring the idea of risk aversion to bear on the determination of the properties of $u_s(\cdot)$. Thus, the approach via the extended independence axiom assumes a stronger basic framework (preferences are defined on the set $\mathscr{L}$ rather than on the smaller $\mathbb{R}^S_+$), but it also yields stronger conclusions.

6.F Subjective Probability Theory

Up to this point in the development of the theory, we have been assuming that risk, summarized by means of numerical probabilities, is regarded as an objective fact by the decision maker. But this is rarely true in reality. Individuals make judgments about the chances of uncertain events that are not necessarily expressible in quantitative form. Even when probabilities are mentioned, as sometimes happens when a doctor discusses the likelihood of various outcomes of medical treatment, they are often acknowledged as imprecise *subjective* estimates.

It would be very helpful, both theoretically and practically, if we could assert that choices are made *as if* individuals held probabilistic beliefs. Even better, we would like that well-defined probabilistic beliefs be revealed by choice behavior. This is the intent of *subjective probability theory*. The theory argues that even if states of the world are not associated with recognizable, objective probabilities, consistency-like restrictions on preferences among gambles still imply that decision makers behave *as if* utilities were assigned to outcomes, probabilities were attached to states of nature, and decisions were made by taking expected utilities. Moreover, this rationalization of the decision maker's behavior with an expected utility function can be done uniquely (up to a positive linear transformation for the utility functions). The theory is therefore a far-reaching generalization of expected utility theory. The classical reference for subjective probability theory is Savage (1954), which is very readable but also advanced. It is, however, possible to gain considerable insight into the theory if one is willing to let the analysis be aided by the use of lotteries with objective random outcomes. This is the approach suggested by Anscombe and Aumann (1963), and we will follow it here.

We begin, as in Section 6.E, with a set of states $\{1, \dots, S\}$. The probabilities on $\{1, \dots, S\}$ are not given. In effect, we aim to *deduce* them. As before, a random variable with monetary payoffs is a vector $x = (x_1, \dots, x_S) \in \mathbb{R}^S_+$.[26] We also want to allow for the possibility that the monetary payoffs in a state are not certain but are themselves money lotteries with objective distributions F_s. Thus, our set of risky alternatives, denoted $\mathscr{L}$, is the set of all S-tuples $(F_1, \dots, F_S)$ of distribution functions.

Suppose now that we are given a rational preference relation $\succsim$ on $\mathscr{L}$. We assume that $\succsim$ satisfies the continuity and the extended independence axioms introduced in Section 6.E. Then, by Proposition 6.E.1, we conclude that there are $u_s(\cdot)$ such that any $(x_1, \dots, x_S) \in \mathbb{R}^S_+$ can be evaluated by $\sum_s u_s(x_s)$. In addition, $u_s(\cdot)$ is a Bernoulli utility function for money lotteries in state s.

The existence of the $u_s(\cdot)$ functions does not yet allow us to identify subjective probabilities. Indeed, for *any* $(\pi_1, \dots, \pi_S) \gg 0$, we could define $\tilde{u}_s(\cdot) = (1/\pi_s)u_s(\cdot)$, and we could then evaluate $(x_1, \dots, x_S)$ by $\sum_s \pi_s \tilde{u}_s(x_s)$. What is needed is some way to disentangle utilities from probabilities.

Consider an example. Suppose that a gamble that gives one dollar in state 1 and none in state 2 is preferred to a gamble that gives one dollar in state 2 and none in state 1. Provided *there is no reason to think that the labels of the states have any*

26. To be specific, we consider monetary payoffs here. All the subsequent arguments, however, work with arbitrary sets of outcomes.

particular influence on the value of money, it is then natural to conclude that the decision maker regards state 2 as less likely than state 1.

This example suggests an additional postulate. Preferences over money lotteries within state s should be the same as those within any other state s'; that is, risk attitudes towards money gambles should be the same across states. To formulate such a property, we define the state s preferences $\succsim_s$ on state s lotteries by

$$F_s \succsim_s F_s' \quad \text{if} \quad \int u_s(x_s)\, dF_s(x_s) \geq \int u_s(x_s)\, dF_s'(x_s).$$

Definition 6.F.1: The state preferences $(\succsim_1, \ldots, \succsim_S)$ on state lotteries are *state uniform* if $\succsim_s = \succsim_{s'}$ for any s and s'.

With state uniformity, $u_s(\cdot)$ and $u_{s'}(\cdot)$ can differ only by an increasing linear transformation. Therefore, there is $u(\cdot)$ such that, for all $s = 1, \ldots, S$,

$$u_s(\cdot) = \pi_s u(\cdot) + \beta_s$$

for some $\pi_s > 0$ and β_s. Moreover, because we still represent the same preferences if we divide all π_s and β_s by a common constant, we can normalize the π_s so that $\sum_s \pi_s = 1$. These π_s are going to be our subjective probabilities.

Proposition 6.F.1: (*Subjective Expected Utility Theorem*) Suppose that the preference relation $\succsim$ on $\mathscr{L}$ satisfies the continuity and extended independence axioms. Suppose, in addition, that the derived state preferences are state uniform. Then there are probabilities $(\pi_1, \ldots, \pi_S) \gg 0$ and a utility function $u(\cdot)$ on amounts of money such that for any $(x_1, \ldots, x_S)$ and $(x_1', \ldots, x_S')$ we have

$$(x_1, \ldots, x_S) \succsim (x_1', \ldots, x_S') \quad \text{if and only if} \quad \sum_s \pi_s u(x_s) \geq \sum_s \pi_s u(x_s').$$

Moreover, the probabilities are uniquely determined, and the utility function is unique up to origin and scale.

Proof: Existence has already been proven. You are asked to establish uniqueness in Exercise 6.F.1. ■

The practical advantages of the subjective expected utility representation are similar to those of the objective version, which we discussed in Section 6.B, and we will not repeat them here. A major virtue of the theory is that it gives a precise, quantifiable, and operational meaning to uncertainty. It is, indeed, most pleasant to be able to remain in the familiar realm of the probability calculus.

But there are also problems. The plausibility of the axioms cannot be completely dissociated from the complexity of the choice situations. The more complex these become, the more strained even seemingly innocent axioms are, For example, is the completeness axiom reasonable for preferences defined on huge sets of random variables? Or consider the implicit axiom (often those are the most treacherous) that the situation can actually be formalized as indicated by the model. This posits the ability to list all conceivable states of the world (or, at least, a sufficiently disaggregated version of this list). In summary, every difficulty so far raised against our model of the rational consumer (i.e., to transitivity, to completeness, to independence) will apply with increased force to the current model.

There are also difficulties specific to the nonobjective nature of probabilities. We devote Example 6.F.1 to this point.

Example 6.F.1: This example is a variation of the *Ellsberg paradox.*[27] There are two urns, denoted R and H. Each urn contains 100 balls. The balls are either white or black. Urn R contains 49 white balls and 51 black balls. Urn H contains an unspecified assortment of balls. A ball has been randomly picked from each urn. Call them the *R-ball* and the *H-ball*, respectively. The color of these balls has not been disclosed. Now we consider two choice situations. In both experiments, the decision maker must choose either the R-ball or the H-ball. After the choices have been made, the color will be disclosed. In the first choice situation, a prize of 1000 dollars is won if the chosen ball is black. In the second choice situation, the same amount is won if the ball is white. With the information given, most people will choose the R-ball in the first experiment. If the decision is made using subjective probabilities, this should mean that the subjective probability that the H-ball is white is larger than .49. Hence, most people should choose the H-ball in the second experiment. However, it turns out that this does not happen overwhelmingly in actual experiments. The decision maker understands that by choosing the R-ball, he has only a 49% chance of winning. However, this chance is "safe" and well understood. The uncertainties incurred are much less clear if he chooses the H-ball. ■

Knight (1921) proposed distinguishing between *risk* and *uncertainty* according to whether the probabilities are given to us objectively or not. In a sense, the theory of subjective probability nullifies this distinction by reducing all uncertainty to risk through the use of beliefs expressible as probabilities. The Example 6.F.1 suggests that there may be something to the distinction. This is an active area of research [e.g., Bewley (1986) and Gilboa and Schmeidler (1989)].

27. From Ellsberg (1961).

REFERENCES

[illegible] (1953). Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Américaine. *Econometrica* **21**: 503–46.

[illegible] F., and R. Aumann. (1963). A definition of subjective probability. *Annals of Mathematical Statistics* **34**: 199–205.

[illegible] K. J. (1971). *Essays in the Theory of Risk Bearing.* Chicago: Markham.

[illegible] T. (1986). Knightian Decision Theory: Part 1. New Haven: Cowles Foundation Discussion Paper No. 807.

[illegible] E. (1986). An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom. *Journal of Economic Theory* **40**: 304–18.

[illegible] P., and M. Rothschild. (1978). *Uncertainty in Economics: Readings and Exercises.* New York: Academic Press.

[illegible] D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* **75**: 643–69.

[illegible] I., and D. Schmeidler. (1989). Maximin expected utility with a unique prior. *Journal of Mathematical Economics* **18**: 141–53.

[illegible] D., and C. H. Plott. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review* **69**: 623–38.

[illegible] J. (1987). 'Making book against oneself,' the independence axiom, and nonlinear utility theory. *Quarterly Journal of Economics* **98**: 785–96.

[illegible] D. and C. Orme. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica* **62**: 1291–326.

Knight, F. (1921). *Risk, Uncertainty and Profit.* Boston, Mass.: Houghton Mifflin. Reprint, London: London School of Economics 1946.

Kreps, D. (1988). *Notes on the Theory of Choice.* Boulder, Colo.: Westview Press.

Machina, M. (1987). Choice under uncertainty: Problems solved and unsolved. *The Journal of Perspectives* **1**: 121–54.

Pratt, J. (1964). Risk aversion in the small and in the large. *Econometrica* **32**: 122–36. Reprinted in Diamond and Rothschild.

Rothschild, M. and J. Stiglitz. (1970). Increasing risk I: A definition. *Journal of Economic Theory* **2**: 225–43. Reprinted in Diamond and Rothschild.

Savage, L. (1954). *The Foundations of Statistics.* New York: Wiley.

Von Neumann, J., and O. Morgenstern. (1944). *Theory of Games and Economic Behavior.* Princeton, N.J.: Princeton University Press.

EXERCISES

6.B.1[A] In text.

6.B.2[A] In text.

6.B.3[B] Show that if the set of outcomes C is finite and the rational preference relation $\succsim$ on the set of lotteries $\mathscr{L}$ satisfies the independence axiom, then there are best and worst lotteries in $\mathscr{L}$. That is, we can find lotteries $\bar{L}$ and $\underline{L}$ such that $\bar{L} \succsim L \succsim \underline{L}$ for all $L \in \mathscr{L}$.

6.B.4[B] The purpose of this exercise is to illustrate how expected utility theory allows us to make consistent decisions when dealing with extremely small probabilities by considering relatively large ones. Suppose that a safety agency is thinking of establishing a criterion under which an area prone to flooding should be evacuated. The probability of flooding is 1%. There are four possible outcomes:

(A) No evacuation is necessary, and none is performed.
(B) An evacuation is performed that is unnecessary.
(C) An evacuation is performed that is necessary.
(D) No evacuation is performed, and a flood causes a disaster.

Suppose that the agency is indifferent between the sure outcome B and the lottery of A with probability p and D with probability $1 - p$, and between the sure outcome C and the lottery of B with probability q and D with probability $1 - q$. Suppose also that it prefers A to D and that $p \in (0, 1)$ and $q \in (0, 1)$. Assume that the conditions of the expected utility theorem are satisfied.

(a) Construct a utility function of the expected utility form for the agency.

(b) Consider two different policy criteria:

Criterion 1: This criterion will result in an evacuation in 90% of the cases in which flooding will occur and an unnecessary evacuation in 10% of the cases in which no flooding occurs.

Criterion 2: This criterion is more conservative. It results in an evacuation in 95% of the cases in which flooding will occur and an unnecessary evacuation in 5% of the cases in which no flooding occurs.

First, derive the probability distributions over the four outcomes under these two criteria. Then, by using the utility function in **(a)**, decide which criterion the agency would prefer.

6.B.5[B] The purpose of this exercise is to show that the Allais paradox is compatible with a weaker version of the independence axiom. We consider the following axiom, known as the

betweenness axiom [see Dekel (1986)]:

For all L, L' and $\lambda \in (0, 1)$, if $L \sim L'$, then $\lambda L + (1 - \lambda)L' \sim L$.

Suppose that there are three possible outcomes.

(a) Show that a preference relation on lotteries satisfying the independence axiom also satisfies the betweenness axiom.

(b) Using a simplex representation for lotteries similar to the one in Figure 6.B.1(b), show that if the continuity and betweenness axioms are satisfied, then the indifference curves of a preference relation on lotteries are straight lines. Conversely, show that if the indifference curves are straight lines, then the betweenness axiom is satisfied. Do these straight lines need to be parallel?

(c) Using **(b)**, show that the betweenness axiom is weaker (less restrictive) than the independence axiom.

(d) Using Figure 6.B.7, show that the choices of the Allais paradox are compatible with the betweenness axiom by exhibiting an indifference map satisfying the betweenness axiom that yields the choices of the Allais paradox.

6.B.6[B] Prove that the induced utility function $U(\cdot)$ defined in the last paragraph of Section 6.B is convex. Give an example of a set of outcomes and a Bernoulli utility function for which the induced utility function is not linear.

6.B.7[A] Consider the following two lotteries:

$$L: \begin{cases} 200 \text{ dollars with probability } .7. \\ 0 \text{ dollars with probability } .3. \end{cases}$$

$$L': \begin{cases} 1200 \text{ dollars with probability } .1. \\ 0 \text{ dollars with probability } .9. \end{cases}$$

Let x_L and $x_{L'}$ be the sure amounts of money that an individual finds indifferent to L and L'. Show that if his preferences are transitive and monotone, the individual must prefer L to L' if and only if $x_L > x_{L'}$. [*Note*: In actual experiments, however, a preference reversal is often observed in which L is preferred to L' but $x_L < x_{L'}$. See Grether and Plott (1979) for details.]

6.C.1[B] Consider the insurance problem studied in Example 6.C.1. Show that if insurance is not actuarially fair (so that $q > \pi$), then the individual will not insure completely.

6.C.2[B]

(a) Show that if an individual has a Bernoulli utility function $u(\cdot)$ with the quadratic form

$$u(x) = \beta x^2 + \gamma x,$$

then his utility from a distribution is determined by the mean and variance of the distribution and, in fact, by these moments alone. [*Note*: The number β should be taken to be negative in order to get the concavity of $u(\cdot)$. Since $u(\cdot)$ is then decreasing at $x > -\gamma/2\beta$, $u(\cdot)$ is useful only when the distribution cannot take values larger than $-\gamma/2\beta$.]

(b) Suppose that a utility function $U(\cdot)$ over distributions is given by

$$U(F) = (\text{mean of } F) - r(\text{variance of } F),$$

where $r > 0$. Argue that unless the set of possible distributions is further restricted (see, e.g., Exercise 6.C.19), $U(\cdot)$ cannot be compatible with any Bernoulli utility function. Give an example of two lotteries L and L' over the same two amounts of money (say x' and $x'' > x'$) such that L gives a higher probability to x'' than does L' and yet according to $U(\cdot)$, L' is preferred to L.

6.C.3$^{\mathbf{B}}$ Prove that the four conditions of Proposition 6.C.1 are equivalent. [*Hint*: The equivalence of (i), (ii), and (iii) has already been shown. As for (iv), prove that (i) implies (iv) and that (iv) implies $u(\frac{1}{2}x + \frac{1}{2}y) \geq \frac{1}{2}u(x) + \frac{1}{2}u(y)$ for any x and y, which is, in fact, sufficient for (ii).]

6.C.4$^{\mathbf{B}}$ Suppose that there are N risky assets whose returns z_n $(n = 1, \ldots, N)$ per dollar invested are jointly distributed according to the distribution function $F(z_1, \ldots, z_N)$. Assume also that all the returns are nonnegative with probability one. Consider an individual who has a continuous, increasing, and concave Bernoulli utility function $u(\cdot)$ over $\mathbb{R}_+$. Define the utility function $U(\cdot)$ of this investor over $\mathbb{R}_+^N$, the set of all nonnegative portfolios, by

$$U(\alpha_1, \ldots, \alpha_N) = \int u(\alpha_1 z_1 + \cdots + \alpha_N z_N)\, dF(z_1, \ldots, z_N).$$

Prove that $U(\cdot)$ is (a) increasing, (b) concave, and (c) continuous (this is harder).

6.C.5$^{\mathbf{A}}$ Consider a decision maker with utility function $u(\cdot)$ defined over $\mathbb{R}_+^L$, just as in Chapter 3.

(a) Argue that concavity of $u(\cdot)$ can be interpreted as the decision maker exhibiting risk aversion with respect to lotteries whose outcomes are bundles of the L commodities.

(b) Suppose now that a Bernoulli utility function $u(\cdot)$ for wealth is derived from the maximization of a utility function defined over bundles of commodities for each given wealth level w, while prices for those commodities are fixed. Show that, if the utility function for the commodities exhibits risk aversion, then so does the derived Bernoulli utility function for wealth. Interpret.

(c) Argue that the converse of part **(b)** does not need to hold: There are nonconcave functions $u\colon \mathbb{R}_+^L \to \mathbb{R}$ such that for any price vector the derived Bernoulli utility function on wealth exhibits risk aversion.

6.C.6$^{\mathbf{B}}$ For Proposition 6.C.2:

(a) Prove the equivalence of conditions (ii) and (iii).

(b) Prove the equivalence of conditions (iii) and (v).

6.C.7$^{\mathbf{A}}$ Prove that, in Proposition 6.C.2, condition (iii) implies condition (iv), and (iv) implies (i).

6.C.8$^{\mathbf{A}}$ In text.

6.C.9$^{\mathbf{B}}$ (M. Kimball) The purpose of this problem is to examine the implications of uncertainty and precaution in a simple consumption–savings decision problem.

In a two-period economy, a consumer has first-period initial wealth w. The consumer's utility level is given by

$$u(c_1, c_2) = u(c_1) + v(c_2),$$

where $u(\cdot)$ and $v(\cdot)$ are concave functions and c_1 and c_2 denote consumption levels in the first and the second period, respectively. Denote by x the amount saved by the consumer in the first period (so that $c_1 = w - x$ and $c_2 = x$), and let x_0 be the optimal value of x in this problem.

We now introduce uncertainty in this economy. If the consumer saves an amount x in the first period, his wealth in the second period is given by $x + y$, where y is distributed according to $F(\cdot)$. In what follows, $E[\cdot]$ always denotes the expectation with respect to $F(\cdot)$. Assume that the Bernoulli utility function over realized wealth levels in the two periods (w_1, w_2) is $u(w_1) + v(w_2)$. Hence, the consumer now solves

$$\operatorname*{Max}_{x}\ u(w - x) + E[v(x + y)].$$

Denote the solution to this problem by x^*.

(a) Show that if $E[v'(x_0 + y)] > v'(x_0)$, then $x^* > x_0$.

(b) Define the *coefficient of absolute prudence* of a utility function $v(\cdot)$ at wealth level x to be $-v'''(x)/v''(x)$. Show that if the coefficient of absolute prudence of a utility function $v_1(\cdot)$ is larger than the coefficient of absolute prudence of utility function $v_2(\cdot)$ for all levels of wealth, then $E[v_1'(x_0 + y)] > v_1'(x_0)$ implies $E[v_2'(x_0 + y)] > v_2'(x_0)$. What are the implications of this fact in the context of part **(a)**?

(c) Show that if $v'''(\cdot) > 0$, and $E[y] = 0$, then $E[v'(x + y)] > v'(x)$ for all values of x.

(d) Show that if the coefficient of absolute risk aversion of $v(\cdot)$ is decreasing with wealth, then $-v'''(x)/v''(x) > -v''(x)/v'(x)$ for all x, and hence $v'''(\cdot) > 0$.

6.C.10 Prove the equivalence of conditions (i) through (v) in Proposition 6.C.3. [*Hint*: By defining $u_1(z) = u(w_1 + z)$ and $u_2(z) = u(w_2 + z)$, show that each of the five conditions in Proposition 6.C.3 is equivalent to the counterpart in Proposition 6.C.2.]

6.C.11[B] For the model in Example 6.C.2, show that if $r_R(x, u)$ is increasing in x then the proportion of wealth invested in the risky asset $\gamma = \alpha/x$ is decreasing with x. Similarly, if $r_R(x, u)$ is decreasing in x, then $\gamma = \alpha/x$ is increasing in x. [*Hint*: Let $u_1(t) = u(tw_1)$ and $u_2(t) = u(tw_2)$, and use the fact, stated in the analysis of Example 6.C.2, that if one Bernoulli utility function is more risk averse than another, then the optimal level of investment in the risky asset for the first function is smaller than that for the second function. You could also give a direct proof using first-order conditions.]

6.C.12[B] Let $u: \mathbb{R}_+ \to \mathbb{R}$ be a strictly increasing Bernoulli utility function. Show that

(a) $u(\cdot)$ exhibits constant relative risk aversion equal to $\rho \neq 1$ if and only if $u(x) = \beta x^{1-\rho} + \gamma$, where $\beta > 0$ and $\gamma \in \mathbb{R}$.

(b) $u(\cdot)$ exhibits constant relative risk aversion equal to 1 if and only if $u(x) = \beta \ln x + \gamma$, where $\beta > 0$ and $\gamma \in \mathbb{R}$.

(c) $\lim_{\rho \to 1} (x^{1-\rho}/(1-\rho)) = \ln x$ for all $x > 0$.

6.C.13[B] Assume that a firm is risk neutral with respect to profits and that if there is any uncertainty in prices, production decisions are made after the resolution of such uncertainty. Suppose that the firm faces a choice between two alternatives. In the first, prices are uncertain. In the second, prices are nonrandom and equal to the expected price vector in the first alternative. Show that a firm that maximizes expected profits will prefer the first alternative over the second.

6.C.14 Consider two risk-averse decision makers (i.e., two decision makers with concave Bernoulli utility functions) choosing among monetary lotteries. Define the utility function $u^*(\cdot)$ to be strongly more risk averse than $u(\cdot)$ if and only if there is a positive constant k and a nonincreasing and concave function $v(\cdot)$ such that $u^*(x) = ku(x) + v(x)$ for all x. The monetary amounts are restricted to lie in the interval $[0, r]$.

(a) Show that if $u^*(\cdot)$ is strongly more risk averse than $u(\cdot)$, then $u^*(\cdot)$ is more risk averse than $u(\cdot)$ in the usual Arrow–Pratt sense.

(b) Show that if $u(\cdot)$ is bounded, then there is no $u^*(\cdot)$ other than $u^*(\cdot) = ku(\cdot) + c$, where c is a constant, that is strongly more risk averse than $u(\cdot)$ on the entire interval $[0, +\infty]$. [Note: In this part, disregard the assumption that the monetary amounts are restricted to lie in the interval $[0, r]$.]

(c) Using (b), argue that the concept of a strongly more risk-averse utility function is stronger (i.e., more restrictive) than the Arrow–Pratt concept of a more risk-averse utility function.

6.C.15[A] Assume that, in a world with uncertainty, there are two assets. The first is a riskless asset that pays 1 dollar. The second pays amounts a and b with probabilities of π and $1 - \pi$, respectively. Denote the demand for the two assets by (x_1, x_2).

Suppose that a decision maker's preferences satisfy the axioms of expected utility theory and that he is a risk averter. The decision maker's wealth is 1, and so are the prices of the assets. Therefore, the decision maker's budget constraint is given by

$$x_1 + x_2 = 1, \quad x_1, x_2 \in [0, 1].$$

(a) Give a simple *necessary* condition (involving a and b only) for the demand for the riskless asset to be strictly positive.

(b) Give a simple *necessary* condition (involving a, b, and π only) for the demand for the risky asset to be strictly positive.

In the next three parts, assume that the conditions obtained in **(a)** and **(b)** are satisfied.

(c) Write down the first-order conditions for utility maximization in this asset demand problem.

(d) Assume that $a < 1$. Show by analyzing the first-order conditions that $dx_1/da \leq 0$.

(e) Which sign do you conjecture for $dx_1/d\pi$? Give an economic interpretation.

(f) Can you prove your conjecture in **(e)** by analyzing the first-order conditions?

6.C.16[A] An individual has Bernoulli utility function $u(\cdot)$ and initial wealth w. Let lottery L offer a payoff of G with probability p and a payoff of B with probability $1 - p$.

(a) If the individual owns the lottery, what is the minimum price he would sell it for?

(b) If he does not own it, what is the maximum price he would be willing to pay for it?

(c) Are buying and selling prices equal? Give an economic interpretation for your answer. Find conditions on the parameters of the problem under which buying and selling prices are equal.

(d) Let $G = 10$, $B = 5$, $w = 10$, and $u(x) = \sqrt{x}$. Compute the buying and selling prices for this lottery and this utility function.

6.C.17[B] Assume that an individual faces a two-period portfolio allocation problem. In period $t = 0, 1$, his wealth w_t is to be divided between a safe asset with return R and a risky asset with return x. The initial wealth at period 0 is w_0. Wealth at period $t = 1, 2$ depends on the portfolio α_{t-1} chosen at period $t - 1$ and on the return x_t realized at period t, according to

$$w_t = ((1 - \alpha_{t-1})R + \alpha_{t-1}x_t)w_{t-1}.$$

The objective of this individual is to maximize the expected utility of terminal wealth w_2. Assume that x_1 and x_2 are independently and identically distributed. Prove that the individual optimally sets $\alpha_0 = \alpha_1$ if his utility function exhibits constant relative risk aversion. Show also that this fails to hold if his utility function exhibits constant absolute risk aversion.

6.C.18[B] Suppose that an individual has a Bernoulli utility function $u(x) = \sqrt{x}$.

(a) Calculate the Arrow–Pratt coefficients of absolute and relative risk aversion at the level of wealth $w = 5$.

(b) Calculate the certainty equivalent and the probability premium for a gamble $(16, 4; \frac{1}{2}, \frac{1}{2})$.

(c) Calculate the certainty equivalent and the probability premium for a gamble $(36, 16; \frac{1}{2}, \frac{1}{2})$. Compare this result with the one in **(b)** and interpret.

6.C.19[C] Suppose that an individual has a Bernoulli utility function $u(x) = -e^{-\alpha x}$ where $\alpha > 0$. His (nonstochastic) initial wealth is given by w. There is one riskless asset and there are N

assets. The return per unit invested on the riskless asset is r. The returns of the risky are jointly normally distributed random variables with means $\mu = (\mu_1, \ldots, \mu_N)$ and ance–covariance matrix V. Assume that there is no redundancy in the risky assets, so that of full rank. Derive the demand function for these $N + 1$ assets.

Consider a lottery over monetary outcomes that pays $x + \varepsilon$ with probability $\frac{1}{2}$ and $-\varepsilon$ with probability $\frac{1}{2}$. Compute the second derivative of this lottery's certainty equivalent respect to ε. Show that the limit of this derivative as $\varepsilon \to 0$ is exactly $-r_A(x)$.

The purpose of this exercise is to prove Proposition 6.D.1 in a two-dimensional ility simplex. Suppose that there are three monetary outcomes: 1 dollar, 2 dollars, and ars. Consider the probability simplex of Figure 6.B.1(b).

(a) For a given lottery L over these outcomes, determine the region of the probability in which lie the lotteries whose distributions first-order stochastically dominate the tion of L.

(b) Given a lottery L, determine the region of the probability simplex in which lie the es L' such that $F(x) \leq G(x)$ for every x, where $F(\cdot)$ is the distribution of L' and $G(\cdot)$ is stribution of L. [Notice that we get the same region as in **(a)**.]

Prove that if $F(\cdot)$ first-order stochastically dominates $G(\cdot)$, then the mean of x under $\int x\, dF(x)$, exceeds that under $G(\cdot)$, $\int x\, dG(x)$. Also provide an example where $\int x\, dF(x) >$ $x)$ but $F(\cdot)$ does not first-order stochastically dominate $G(\cdot)$.

Verify that if a distribution $G(\cdot)$ is an elementary increase in risk from a distribution then $F(\cdot)$ second-order stochastically dominates $G(\cdot)$.

The purpose of this exercise is to verify the equivalence of the three statements of sition 6.D.2 in a two-dimensional probability simplex. Suppose that there are three ary outcomes: 1, 2, and 3 dollars. Consider the probability simplex in Figure 6.B.1(b).

(a) If two lotteries have the same mean, what are their positions relative to each other in probability simplex.

(b) Given a lottery L, determine the region of the simplex in which lie the lotteries L' distributions are second-order stochastically dominated by the distribution of L.

(c) Given a lottery L, determine the region of the simplex in which lie the lotteries L' whose tions are mean preserving spreads of L.

(d) Given a lottery L, determine the region of the simplex in which lie the lotteries L' such condition (6.D.2) holds, where $F(\cdot)$ and $G(\cdot)$ are, respectively, the distributions of

ce that in **(b)**, **(c)**, and **(d)**, you always have the same region.

The purpose of this exercise is to show that preferences may not be transitive in the nce of regret. Let there be S states of the world, indexed by $s = 1, \ldots, S$. Assume that s occurs with probability π_s. Define the expected regret associated with lottery $\ldots, x_S)$ relative to lottery $x' = (x'_1, \ldots, x'_S)$ by

$$\sum_{s=1}^{S} \pi_s h(\text{Max}\,\{0, x'_s - x_s\}),$$

$\cdot)$ is a given increasing function. [We call $h(\cdot)$ the *regret valuation function*; it measures ret the individual has after the state of nature is known.] We define x to be at least as as x' in the presence of regret if and only if the expected regret associated with x relative is not greater than the expected regret associated with x' relative to x.

Suppose that $S = 3$, $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$, and $h(x) = \sqrt{x}$. Consider the following three lotteries:

$$x = (0, -2, \quad 1),$$

$$x' = (0, \quad 2, -2),$$

$$x'' = (2, -3, -1).$$

Show that the preference ordering over these three lotteries is not transitive.

6.E.2^A Assume that in a world with uncertainty there are two possible states of nature ($s = 1, 2$) and a single consumption good. There is a single decision maker whose preferences over lotteries satisfy the axioms of expected utility theory and who is a risk averter. For simplicity, we assume that utility is state-independent.

Two contingent commodities are available to the decision maker. The first (respectively, the second) pays one unit of the consumption good in state $s = 1$ (respectively $s = 2$) and zero otherwise. Denote the vector quantities of the two contingent commodities by (x_1, x_2).

(a) Show that the preference relation of the decision maker on (x_1, x_2) is convex.

(b) Argue that the decision maker is also a risk averter when choosing between lotteries whose outcomes are vectors (x_1, x_2).

(c) Show that the Walrasian demand functions for x_1 and x_2 are normal.

6.E.3^B Let $g: S \to \mathbb{R}_+$ be a random variable with mean $E(g) = 1$. For $\alpha \in (0, 1)$, define a new random variable $g^*: S \to \mathbb{R}_+$ by $g^*(s) = \alpha g(s) + (1 - \alpha)$. Note that $E(g^*) = 1$. Denote by $G(\cdot)$ and $G^*(\cdot)$ the distribution functions of $g(\cdot)$ and $g^*(\cdot)$, respectively. Show that $G^*(\cdot)$ second-order stochastically dominates $G(\cdot)$. Interpret.

6.F.1^B Prove that in the subjective expected utility theorem (Proposition 6.F.2), the obtained utility function $u(\cdot)$ on money is uniquely determined up to origin and scale. That is, if both $u(\cdot)$ and $\hat{u}(\cdot)$ satisfy the condition of the theorem, then there exist $\beta > 0$ and $\gamma \in \mathbb{R}$ such that $\hat{u}(x) = \beta u(x) + \gamma$ for all x. Prove also that the subjective probabilities are uniquely determined.

6.F.2^A The purpose of this exercise is to explain the outcomes of the experiments described in Example 6.F.1 by means of the theory of *nonunique prior beliefs* of Gilboa and Schmeidler (1989).

We consider a decision maker with a Bernoulli utility function $u(\cdot)$ defined on $\{0, 1000\}$. We normalize $u(\cdot)$ so that $u(0) = 0$ and $u(1000) = 1$.

The probabilistic belief that the decision maker might have on the color of the H-ball being white is a number $\pi \in [0, 1]$. We assume that the decision maker has, not a single belief but a *set* of beliefs given by a subset P of $[0, 1]$. The actions that he may take are denoted R or H with R meaning that he chooses the R-ball and H meaning that he chooses the H-ball.

As in Example 6.F.1, the decision maker is faced with two different choice situations. In choice situation W, he receives 1000 dollars if the ball chosen is white and 0 dollars otherwise. In choice situation B, he receives 1000 dollars if the ball chosen is black and 0 dollars otherwise.

For each of the two choice situations, define his utility function over the actions R and H in the following way:

For situation W, U_W: $\{\text{R}, \text{H}\} \to \mathbb{R}$ is defined by

$$U_W(\text{R}) = .49 \quad \text{and} \quad U_W(\text{H}) = \text{Min}\,\{\pi\colon \pi \in P\}.$$

For situation B, U_B: $\{\text{R}, \text{H}\} \to \mathbb{R}$ is defined by

$$U_B(\text{R}) = .51 \quad \text{and} \quad U_B(\text{H}) = \text{Min}\,\{(1 - \pi)\colon \pi \in P\}.$$

y, his utility from choice R is the expected utility of 1000 dollars with the (objective) ility calculated from the number of white and black balls in urn R. However, his utility choice H is the expected utility of 1000 dollars with the probability associated with the pessimistic belief in P.

(a) Prove that if P consists of only one belief, then U_W and U_B are derived from Neumann–Morgenstern utility function and that $U_W(\mathrm{R}) > U_W(\mathrm{H})$ if and only if) $< U_B(\mathrm{H})$.

(b) Find a set P for which $U_W(\mathrm{R}) > U_W(\mathrm{H})$ and $U_B(\mathrm{R}) > U_B(\mathrm{H})$.

Game Theory

In Part I, we analyzed individual decision making, both in abstract decision ... and in more specific economic settings. Our primary aim was to lay the ...work for the study of how the simultaneous behavior of many self-interested ...duals (including firms) generates economic outcomes in market economies. ... the remainder of the book is devoted to this task. In Part II, however, we ... in a more general way how multiperson interactions can be modeled.

A central feature of multiperson interaction is the potential for the presence of ... *interdependence*. In our study of individual decision making in Part I, the ... maker faced situations in which her well-being depended only on the choices ... (possibly with some randomness). In contrast, in multiperson situations ...ategic interdependence, each agent recognizes that the payoff she receives (in ... profits) depends not only on her own actions but also on the actions of ...viduals. The actions that are best for her to take may depend on actions ...er individuals have already taken, on those she expects them to be taking ... time, and even on future actions that they may take, or decide not to ... as a result of her current actions.

The tool that we use for analyzing settings with strategic interdependence is ...ative *game theory*. Although the term "game" may seem to undersell the ... importance, it correctly highlights the theory's central feature: The agents ... are concerned with strategy and winning (in the general sense of utility ...maximization) in much the same way that players of most parlor games are.

...person economic situations vary greatly in the degree to which strategic ... is present. In settings of monopoly (where a good is sold by only a single ... Section 12.B) or of perfect competition (where all agents act as price takers; ... 10 and Part IV), the nature of strategic interaction is minimal enough ... analysis need not make any formal use of game theory.[1] In other settings, ... such as the analysis of oligopolistic markets (where there is more than one

1. However, we could well do so in both cases; see, for example, the proof of existence of ... equilibrium in Chapter 17, Appendix B. Moreover, we shall stress how perfect ... can be viewed usefully as a limiting case of oligopolistic strategic interaction; see, for ... Section 12.F.

but still not many sellers of a good; see Sections 12.C to 12.G), the central role of strategic interaction makes game theory indispensable for our analysis.

Part II is divided into three chapters. Chapter 7 provides a short introduction to the basic elements of noncooperative game theory, including a discussion of exactly what a game is, some ways of representing games, and an introduction to a central concept of the theory, a player's *strategy*. Chapter 8 addresses how we can predict outcomes in the special class of games in which all the players move simultaneously, known as *simultaneous-move games*. This restricted focus helps us isolate some central issues while deferring a number of more difficult ones. Chapter 9 studies *dynamic games* in which players' moves may precede one another, and in which some of these more difficult (but also interesting) issues arise.

Note that we have used the modifier *noncooperative* to describe the type of game theory we discuss in Part II. There is another branch of game theory, known as *cooperative game theory*, that we do not discuss here. In contrast with noncooperative game theory, the fundamental units of analysis in cooperative theory are groups and subgroups of individuals that are assumed, as a primitive of the theory, to be able to attain particular outcomes for themselves through binding cooperative agreements. Cooperative game theory has played an important role in general equilibrium theory, and we provide a brief introduction to it in Appendix A of Chapter 18. We should emphasize that the term *noncooperative game theory* does *not* mean that noncooperative theory is incapable of explaining cooperation within groups of individuals. Rather, it focuses on how cooperation may emerge as rational behavior in the absence of an ability to make binding agreements (e.g., see the discussion of repeated interaction among oligopolists in Chapter 12).

Some excellent recent references for further study of noncooperative game theory are Fudenberg and Tirole (1991), Myerson (1992), and Osborne and Rubinstein (1994), and at a more introductory level Gibbons (1992) and Binmore (1992). Kreps (1990) provides a very interesting discussion of some of the strengths and weaknesses of the theory. Von Neumann and Morgenstern (1944), Luce and Raiffa (1957), and Schelling (1960) remain classic references.

REFERENCES

Binmore, K. (1992). *Fun and Games: A Text on Game Theory*. Lexington, Mass.: D. C. Heath.
Fudenberg, D., and J. Tirole. (1991). *Game Theory*. Cambridge, Mass.: MIT Press.
Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton, N.J.: Princeton University Press.
Kreps, D. M. (1990). *Game Theory and Economic Modeling*. New York: Oxford University Press.
Luce, R. D., and H. Raiffa. (1957). *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.
Myerson, R. B. (1992). *Game Theory: Analysis of Conflict*. Cambridge, Mass.: Harvard University Press.
Osborne, M. J., and A. Rubinstein. (1994). *A Course in Game Theory*, Cambridge, Mass.: MIT Press.
Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
Von Neumann, J., and O. Morgenstern. (1944). *The Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.

CHAPTER

7

Basic Elements of Noncooperative Games

7.A Introduction

In this chapter, we begin our study of noncooperative game theory by introducing some of its basic building blocks. This material serves as a prelude to our analysis of games in Chapters 8 and 9.

Section 7.B begins with an informal introduction to the concept of a *game*. It describes the four basic elements of any setting of strategic interaction that we must know to specify a game.

In Section 7.C, we show how a game can be described by means of what is called the *extensive form representation*. The extensive form representation provides a very rich description of a game, capturing who moves when, what they can do, what they know when it is their turn to move, and the outcomes associated with any collection of actions taken by the individuals playing the game.

In Section 7.D, we introduce a central concept of game theory, a player's *strategy*. A player's strategy is a complete contingent plan describing the actions she will take in each conceivable evolution of the game. We then show how the notion of a strategy can be used to derive a much more compact representation of a game, known as the *normal* (or *strategic*) *form representation*.

In Section 7.E, we consider the possibility that a player might randomize her choices. This gives rise to the notion of a *mixed strategy*.

7.B What Is a Game?

A *game* is a formal representation of a situation in which a number of individuals interact in a setting of *strategic interdependence*. By that, we mean that each individual's welfare depends not only on her own actions but also on the actions of the other individuals. Moreover, the actions that are best for her to take may depend on what she expects the other players to do.

To describe a situation of strategic interaction, we need to know four things:

(i) *The players:* Who is involved?

(ii) *The rules:* Who moves when? What do they know when they move? What can they do?

(iii) *The outcomes:* For each possible set of actions by the players, what is the outcome of the game?

(iv) *The payoffs:* What are the players' preferences (i.e., utility functions) over the possible outcomes?

We begin by considering items (i) to (iii). A simple example is provided by the school-yard game of *Matching Pennies.*

Example 7.B.1: *Matching Pennies.* Items (i) to (iii) are as follows:

Players: There are two players, denoted 1 and 2.

Rules: Each player simultaneously puts a penny down, either heads up or tails up.

Outcomes: If the two pennies match (either both heads up or both tails up), player 1 pays 1 dollar to player 2; otherwise, player 2 pays 1 dollar to player 1. ■

Consider another example, the game of *Tick-Tack-Toe.*

Example 7.B.2: *Tick-Tack-Toe.* Items (i) to (iii) are as follows:

Players: There are two players, X and O.

Rules: The players are faced with a board that consists of nine squares arrayed with three rows of three squares each stacked on one another (see Figure 7.B.1). The players take turns putting their marks (an X or an O) into an as-yet-unmarked square. Player X moves first. Both players observe all choices previously made.

Outcomes: The first player to have three of her marks in a row (horizontally, vertically, or diagonally) wins and receives 1 dollar from the other player. If no one succeeds in doing so after all nine boxes are marked, the game is a tie and no payments are made or received by either player. ■

To complete our description of these two games, we need to say what the players' preferences are over the possible outcomes [item (iv) in our list]. As a general matter, we describe a player's preferences by a utility function that assigns a utility level for each possible outcome. It is common to refer to the player's utility function as her *payoff function* and the utility level as her *payoff.* Throughout, we assume that these utility functions take an expected utility form (see Chapter 6) so that when we consider situations in which outcomes are random, we can evaluate the random prospect by means of the player's expected utility.

Figure 7.B.1
A Tick-T[illegible]
board.

In later references to Matching Pennies and Tick-Tack-Toe, we assume that each player's payoff is simply equal to the amount of money she gains or loses. Note that in both examples, the actions that maximize a player's payoff depend on what she expects her opponent to do.

Examples 7.B.1 and 7.B.2 involve situations of pure conflict: What one player wins the other player loses. Such games are called *zero-sum games*. But strategic interaction and game theory are not limited to situations of pure or even partial conflict. Consider the situation in Example 7.B.3.

Example 7.B.3: *Meeting in New York.* Items (i) to (iv) are as follows:

Players: Two players, Mr. Thomas and Mr. Schelling.

Rules: The two players are separated and cannot communicate. They are supposed to meet in New York City at noon for lunch but have forgotten to specify where. Each must decide where to go (each can make only one choice).

Outcomes: If they meet each other, they get to enjoy each other's company at lunch. Otherwise, they must eat alone.

Payoffs: They each attach a monetary value of 100 dollars to the other's company (their payoffs are each 100 dollars if they meet, 0 dollars if they do not).

In this example, the two players' interests are completely aligned. Their problem is simply one of coordination. Nevertheless, each player's payoff depends on what the other player does; and more importantly, *each player's optimal action depends on what he thinks the other will do.* Thus, even the task of coordination can have a strategic nature. ■

Although the information given in items (i) to (iv) fully describe a game, it is useful for purposes of analysis to represent this information in particular ways. We examine one of these ways in Section 7.C.

7.C The Extensive Form Representation of a Game

If we know the items (i) to (iv) described in Section 7.B (the players, the rules, the outcomes, and the payoffs), then we can formally represent the game in what is known as its *extensive form*. The extensive form captures who moves when, what actions each player can take, what players know when they move, what the outcome is as a function of the actions taken by the players, and the players' payoffs from each possible outcome.

We begin by informally introducing the elements of the extensive form representation through a series of examples. After doing so, we then provide a formal specification of the extensive form (some readers may want to begin with this and then return to the examples).

The extensive form relies on the conceptual apparatus known as a *game tree*. As a starting point, it is useful to begin with a very simple variation of Matching Pennies, which we call *Matching Pennies Version B*.

Example 7.C.1: *Matching Pennies Version B and Its Extensive Form.* Matching Pennies Version B is identical to Matching Pennies (see Example 7.B.1) except

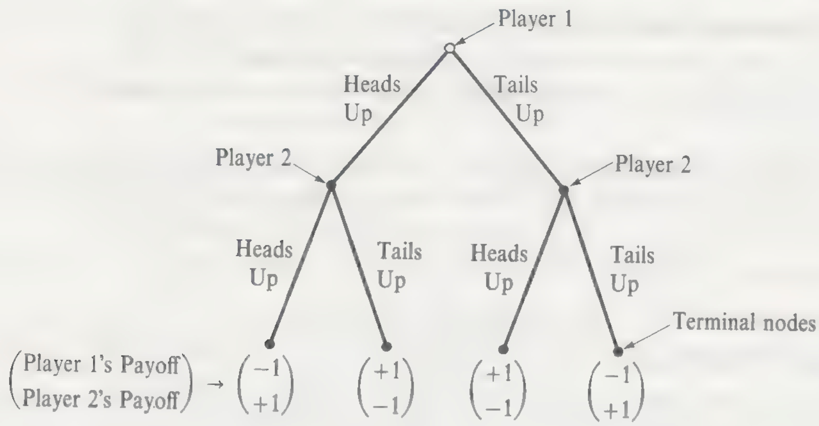


Figure 7.C.1 Extensive form [of] Matching P[ennies] Version B.

that the two players move sequentially, rather than simultaneously. In particular, player 1 puts her penny down (heads up or tails up) first. Then, after seeing player 1's choice, player 2 puts her penny down. (This is a very nice game for player 2!)

The extensive form representation of this game is depicted in Figure 7.C.1. The game starts at an *initial decision node* (represented by an open circle), where player 1 makes her move, deciding whether to place her penny heads up or tails up. Each of the two possible choices for player 1 is represented by a *branch* from this initial decision node. At the end of each branch is another decision node (represented by a solid dot), at which player 2 can choose between two actions, heads up or tails up, after seeing player 1's choice. The initial decision node is referred to as *player 1's decision node*; the latter two as *player 2's decision nodes.* After player 2's move, we reach the end of the game, represented by *terminal nodes.* At each terminal node, we list the players' payoffs arising from the sequence of moves leading to that terminal node.

Note the treelike structure of Figure 7.C.1: Like an actual tree, it has a unique connected path of branches from the initial node (sometimes also called the *root*) to each point in the tree. This type of figure is known as a *game tree.* ■

Example 7.C.2: *The Extensive Form of Tick-Tack-Toe.* The more elaborate game tree shown in Figure 7.C.2 depicts the extensive form for Tick-Tack-Toe (to conserve space, many parts are omitted). Note that every path through the tree represents a unique sequence of moves by the players. In particular, when a given board position (such as the two left corners filled by X and the two right corners filled by O) can be reached through several different sequences of moves, each of these sequences is depicted separately in the game tree. Nodes represent not only the current position but also *how it was reached.* ■

In both Matching Pennies Version B and Tick-Tack-Toe, when it is a player's turn to move, she is able to observe all her rival's previous moves. They are games of *perfect information* (we give a precise definition of this term in Definition 7.C.1). The concept of an *information set* allows us to accommodate the possibility that this is not so. Formally, the elements of an information set are a subset of a particular player's decision nodes. The interpretation is that when play has reached one of the decision nodes in the information set and it is that player's turn to move, she does

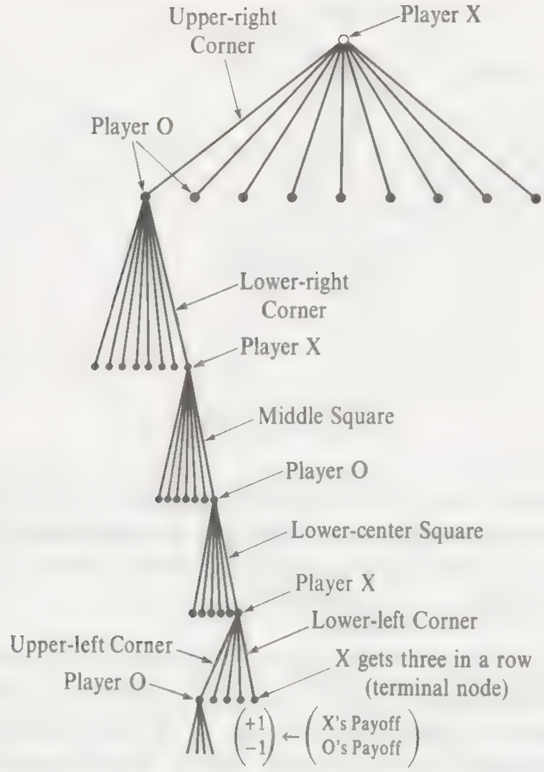


Figure 7.C.2
Part of the extensive form for Tick-Tack-Toe.

does not know which of these nodes she is actually at. The reason for this ignorance is that the player does not observe something about what has previously transpired in the game. A further variation of Matching Pennies, which we call *Matching Pennies Version C*, helps make this concept clearer.

Example 7.C.3: *Matching Pennies Version C and Its Extensive Form.* This version of Matching Pennies is just like Matching Pennies Version B (in Example 7.C.1) except that after player 1 puts her penny down, she keeps it covered with her hand. Hence, player 2 cannot see player 1's choice until after player 2 has moved.

The extensive form for this game is represented in Figure 7.C.3. It is identical to Figure 7.C.1 except that we have drawn a circle around player 2's two decision nodes to indicate that these two nodes are in a single information set. The meaning of this information set is that when it is player 2's turn to move, she cannot tell which of these two nodes she is at because she has not observed player 1's previous move. Note that player 2 has the same two possible actions at each of the two nodes in her information set. This must be the case if player 2 is unable to distinguish the two nodes; otherwise, she could figure out which move player 1 had taken simply by seeing what her own possible actions are.

In principle, we could also associate player 1's decision node with an information set. Because player 1 knows that nothing has happened before it is her turn to move, this information set has only one member (player 1 knows exactly which node she is at when she moves). To be fully rigorous, we should therefore also draw an information set around player 1's decision node in Figure 7.C.3. It is common, however, to

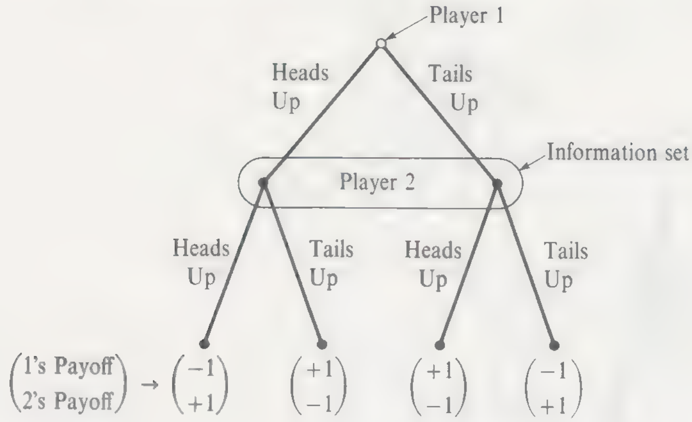


Figure 7.C.3
Extensive
Matching
Version C.

simplify the diagrammatic depiction of a game in extensive form by not drawing the information sets that contain a single node. Thus, any uncircled decision nodes are understood to be elements of *singleton* information sets. In Figures 7.C.1 and 7.C.2, for example, every decision node belongs to a singleton information set. ■

A listing of all of a player's information sets gives a listing, from the player's perspective, of all of the possible distinguishable "events" or "circumstances" in which she might be called upon to move. For example, in Example 7.C.1, from player 2's perspective there are two distinguishable events that might arise in which she would be called upon to move, each one corresponding to play having reached one of her two (singleton) information sets. By way of contrast, player 2 foresees only one possible circumstance in which she would need to move in Example 7.C.3 (this circumstance is, however, certain to arise).

In Example 7.C.3, we noted a natural restriction on information sets: At every node within a given information set, a player must have the same set of possible actions. Another restriction we impose is that players possess what is known as *perfect recall*. Loosely speaking, perfect recall means that a player does not forget what she once knew, including her own actions. Figure 7.C.4 depicts two games in which this condition is not met. In Figure 7.C.4(a), as the game progresses, player 2 forgets a move by player 1 that she once knew (namely, whether player 1 chose ℓ or r). In Figure 7.C.4(b), player 1 forgets her own previous move.[1] All the games we consider in this book satisfy the property of perfect recall.

The use of information sets also allows us to capture play that is simultaneous rather than sequential. This is illustrated in Example 7.C.4 for the game of (standard) Matching Pennies introduced in Example 7.B.1.

---

1. In terms of the formal specification of the extensive form given later in this section, if we denote the information set containing decision node x by $H(x)$, a game is formally characterized as one of perfect recall if the following two conditions hold: (i) If $H(x) = H(x')$, x is neither a predecessor nor a successor of x'; and (ii) if x and x' are two decision nodes for player i with $H(x) = H(x')$, and if x'' is a predecessor of x (not necessarily an immediate one) that is also in one of player i's information sets, with a'' being the action at $H(x'')$ on the path to x, then there must be a predecessor node to x' that is an element of $H(x'')$ and the action at this predecessor node that is on the path to x' must also be a''.

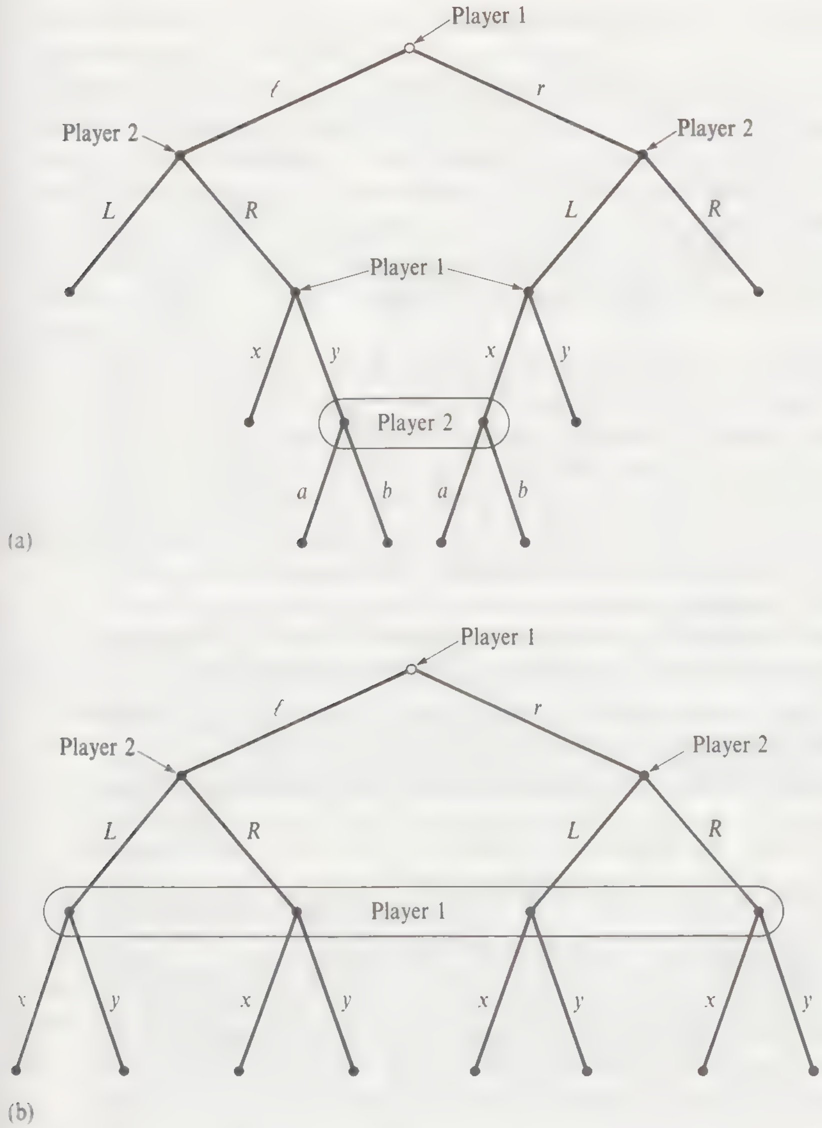


Figure 7.C.4
Two games not satisfying perfect recall.

Example 7.C.4: *The Extensive Form for Matching Pennies.* Suppose now that the players put their pennies down simultaneously. For each player, this game is strategically equivalent to the Version C game. In Version C, player 1 was unable to observe player 2's choice because player 1 moved first, and player 2 was unable to observe player 1's choice because player 1 kept it covered; here each player is unable to observe the other's choice because they move simultaneously. As long as they cannot observe each other's choices, the timing of moves is irrelevant. Thus, we can use the game tree in Figure 7.C.3 to describe the game of (standard) Matching Pennies. Note that by this logic we can also describe this game with a game tree that reverses the decision nodes of players 1 and 2 in Figure 7.C.3. ■

We can now return to the notion of a game of perfect information and offer a formal definition.

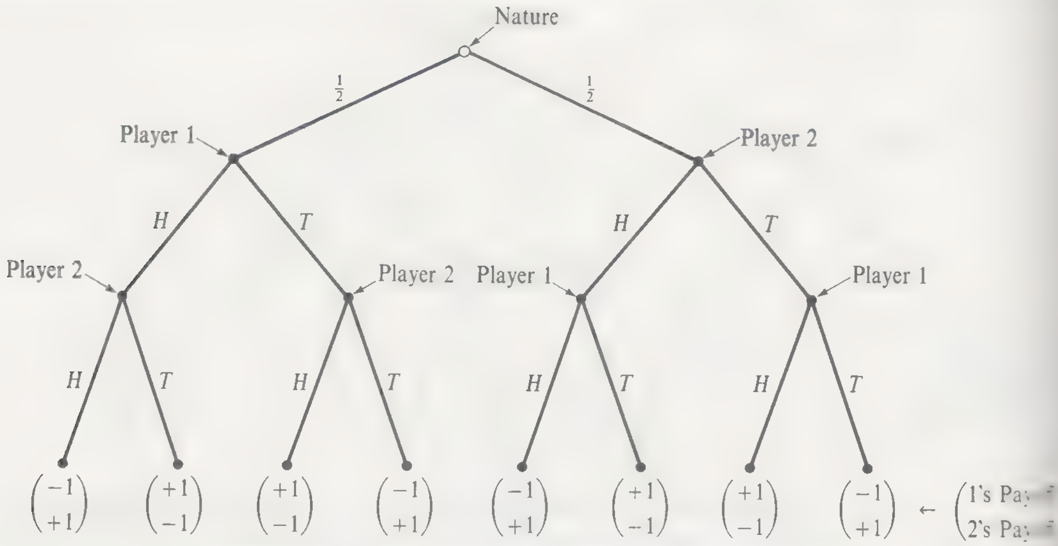


Figure 7.C.5 Extensive form for Matching Pennies Version D.

Definition 7.C.1: A game is one of *perfect information* if each information set contains a single decision node. Otherwise, it is a game of *imperfect information*.

Up to this point, the outcome of a game has been a deterministic function of the players' choices. In many games, however, there is an element of chance. This, too, can be captured in the extensive form representation by including *random moves of nature*. We illustrate this point with still another variation, *Matching Pennies Version D*.

Example 7.C.5: *Matching Pennies Version D and Its Extensive Form.* Suppose that prior to playing Matching Pennies Version B, the two players flip a coin to see who will move first. Thus, with equal probability either player 1 will put her penny down first, or player 2 will. In Figure 7.C.5, this game is depicted as beginning with a *move of nature* at the initial node that has two branches, each with probability $\frac{1}{2}$. Note that this is drawn as if nature were an additional player who must play its two actions with fixed probabilities. (In the figure, H stands for "heads up" and T stands for "tails up".) ■

It is a basic postulate of game theory that all players know the structure of the game, know that their rivals know it, know that their rivals know that they know it, and so on. In theoretical parlance, we say that the structure of the game is *common knowledge* [see Aumann (1976) and Milgrom (1981) for discussions of this concept].

In addition to being depicted graphically, the extensive form can be described mathematically. The basic components are fairly easily explained and can help you keep in mind the fundamental building blocks of a game. Formally, a game represented in extensive form consists of the following items:[2]

2. To be a bit more precise about terminology: A collection of items (i) to (vi) is formally known as an extensive *game form*; adding item (vii), the players' preferences over the outcomes, leads to a *game* represented in extensive form. We will not make anything of this distinction here. See Kuhn (1953) or Section 2 of Kreps and Wilson (1982) for additional discussion of this and other points regarding the extensive form.

(i) A finite set of nodes $\mathscr{X}$, a finite set of possible actions $\mathscr{A}$, and a finite set of players $\{1, \ldots, I\}$.

(ii) A function $p: \mathscr{X} \to \{\mathscr{X} \cup \varnothing\}$ specifying a single immediate predecessor of each node x; $p(x)$ is nonempty for all $x \in \mathscr{X}$ but one, designated as the *initial node* x_0. The immediate successor nodes of x are then $s(x) = p^{-1}(x)$, and the set of *all* predecessors and *all* successors of node x can be found by iterating $p(x)$ and $s(x)$. To have a tree structure, we require that these sets be disjoint (a predecessor of node x cannot also be a successor to it). The set of *terminal nodes* is $T = \{x \in \mathscr{X}: s(x) = \varnothing\}$. All other nodes $\mathscr{X} \backslash T$ are known as *decision nodes*.

(iii) A function $\alpha: \mathscr{X} \backslash \{x_0\} \to \mathscr{A}$ giving the action that leads to any noninitial node x from its immediate predecessor $p(x)$ and satisfying the property that if $x', x'' \in s(x)$ and $x' \neq x''$, then $\alpha(x') \neq \alpha(x'')$. The set of choices available at decision node x is $c(x) = \{a \in \mathscr{A}: a = \alpha(x') \text{ for some } x' \in s(x)\}$.

(iv) A collection of information sets $\mathscr{H}$, and a function $H: \mathscr{X} \to \mathscr{H}$ assigning each decision node x to an information set $H(x) \in \mathscr{H}$. Thus, the information sets in $\mathscr{H}$ form a partition of $\mathscr{X}$. We require that all decision nodes assigned to a single information set have the same choices available; formally, $c(x) = c(x')$ if $H(x) = H(x')$. We can therefore write the choices available at information set H as $C(H) = \{a \in \mathscr{A}: a \in c(x) \text{ for } x \in H\}$.

(v) A function $\iota: \mathscr{H} \to \{0, 1, \ldots, I\}$ assigning each information set in $\mathscr{H}$ to the player (or to nature: formally, player 0) who moves at the decision nodes in that set. We can denote the collection of player i's information sets by $\mathscr{H}_i = \{H \in \mathscr{H}: i = \iota(H)\}$.

(vi) A function $\rho: \mathscr{H}_0 \times \mathscr{A} \to [0, 1]$ assigning probabilities to actions at information sets where nature moves and satisfying $\rho(H, a) = 0$ if $a \notin C(H)$ and $\sum_{a \in C(H)} \rho(H, a) = 1$ for all $H \in \mathscr{H}_0$.

(vii) A collection of payoff functions $u = \{u_1(\cdot), \ldots, u_I(\cdot)\}$ assigning utilities to the players for each terminal node that can be reached, $u_i: T \to \mathbb{R}$. As we noted in Section 7.B, because we want to allow for a random realization of outcomes we take each $u_i(\cdot)$ to be a Bernoulli utility function.

Thus, formally, a game in extensive form is specified by the collection $\Gamma_E = \{\mathscr{X}, \mathscr{A}, I, p(\cdot), \alpha(\cdot), \mathscr{H}, H(\cdot), \iota(\cdot), \rho(\cdot), u\}$.

We should note that there are three implicit types of finiteness hidden in the formulation just presented. Because we will often encounter games not sharing these features in the economic applications discussed in later chapters, we briefly identify them here, although without any formal treatment. The formal definition of an extensive form representation of a game can be extended to these infinite cases without much difficulty, although there can be important differences in the predicted outcomes of finite and infinite economic models, as we shall see later (e.g., in Chapters 12 and 20).

First, we have assumed that players have a finite number of actions available at each decision node. This would rule out a game in which, say, a player can choose any number from some interval $[a, b] \subset \mathbb{R}$. In fact, allowing for an infinite set of actions requires that we allow for an infinite set of nodes as well. But with this change, items (i) to (vii) remain the basic elements of an extensive form representation (e.g., decision nodes and terminal nodes are still associated with a unique path through the tree).

Second, we have described the extensive form of a game that must end after a finite number of moves (because the set of decision nodes is finite). Indeed, all the examples we have considered so far fall into this category. There are, however, other types of games. For example, suppose that two players with infinite life spans (perhaps two firms) play Matching Pennies repeatedly every January 1. The players discount the money gained or lost at future dates with interest rate r and seek to maximize their discounted net gains. In this game, there are no terminal nodes. Even so, we can still associate discounted payoffs for the two players with every (infinite) sequence of moves the players make. Of course, actually drawing a complete game tree would be impossible, but the basic elements of the extensive form can nonetheless be captured as before (with payoffs being associated with paths through the tree rather than with terminal nodes).

Third, we may at times also imagine that there are an infinite number of players who take actions in a game. For example, models involving overlapping generations of players (as in various macroeconomic models) have this feature, as do models of entry in which we want to allow for an infinite number of potential firms. In the games of this type that we consider, this issue can be handled in a simple and natural manner.

Note that all three of these extensions require that we relax the assumption that there is a finite set of nodes. Games with a finite number of nodes, such as those we have been considering, are known as *finite games.*

For pedagogical purposes, we restrict our attention in Part II to finite games except where specifically indicated otherwise. The extension of the formal concepts we discuss here to the economic games studied later in the book that do not share these finiteness properties is straightforward.

7.D Strategies and the Normal Form Representation of a Game

A central concept of game theory is the notion of a player's *strategy*. A strategy is a *complete contingent plan*, or *decision rule*, that specifies how the player will act in *every possible distinguishable circumstance* in which she might be called upon to move. Recall that, from a player's perspective, the set of such circumstances is represented by her collection of information sets, with each information set representing a different distinguishable circumstance in which she may need to move (see Section 7.C). Thus, a player's strategy amounts to a specification of how she plans to move at each one of her information sets, should it be reached during play of the game. This is stated formally in Definition 7.D.1.

Definition 7.D.1: Let $\mathscr{H}_i$ denote the collection of player i's information sets, $\mathscr{A}$ the set of possible actions in the game, and $C(H) \subset \mathscr{A}$ the set of actions possible at information set H. A *strategy* for player i is a function $s_i: \mathscr{H}_i \to \mathscr{A}$ such that $s_i(H) \in C(H)$ for all $H \in \mathscr{H}_i$.

The fact that a strategy is a complete contingent plan cannot be overemphasized, and it is often a source of confusion to those new to game theory. When a player specifies her strategy, it is as if she had to write down an instruction book prior to play so that a representative could act on her behalf merely by consulting that book.

As a complete contingent plan, a strategy often specifies actions for a player at information sets that may not be reached during the actual play of the game.

For example, in Tick-Tack-Toe, player O's strategy describes what she will do on her first move if player X starts the game by marking the center square. But in the actual play of the game, player X might not begin in the center; she may instead mark the lower-right corner first, making this part of player O's plan no longer relevant.

In fact, there is an even subtler point: A player's strategy may include plans for actions that her *own* strategy makes irrelevant. For example, a complete contingent plan for player X in Tick-Tack-Toe includes a description of what she will do after she plays "center" and player O then plays "lower-right corner," even though her own strategy may call for her first move to be "upper-left corner." This probably seems strange; its importance will become apparent only when we talk about dynamic games in Chapter 9. Nevertheless, remember: *A strategy is a complete contingent plan that says what a player will do at each of her information sets if she is called on to play there.*

It is worthwhile to consider what the players' possible strategies are for some of the simple Matching Pennies games.

Example 7.D.1: *Strategies in Matching Pennies Version B.* In Matching Pennies Version B, a strategy for player 1 simply specifies her move at the game's initial node. She has two possible strategies: She can play heads (H) or tails (T). A strategy for player 2, on the other hand, specifies how she will play (H or T) at each of her two information sets, that is, how she will play if player 1 picks H *and* how she will play if player 1 picks T. Thus, player 2 has *four* possible strategies.

Strategy 1 (s_1): Play H if player 1 plays H; play H if player 1 plays T.
Strategy 2 (s_2): Play H if player 1 plays H; play T if player 1 plays T.
Strategy 3 (s_3): Play T if player 1 plays H; play H if player 1 plays T.
Strategy 4 (s_4): Play T if player 1 plays H; play T if player 1 plays T. ■

Example 7.D.2: *Strategies in Matching Pennies Version C.* In Matching Pennies Version C, player 1's strategies are exactly the same as in Version B; but player 2 now only has two possible strategies, "play H" and "play T", because she now has only one information set. She can no longer condition her action on player 1's previous action. ■

We will often find it convenient to represent a profile of players' strategy choices in an I-player game by a vector $s = (s_1, \dots, s_I)$, where s_i is the strategy chosen by player i. We will also sometimes write the strategy profile s as (s_i, s_{-i}), where s_{-i} is the $(I-1)$ vector of strategies for players other than i.

The Normal Form Representation of a Game

Every profile of strategies for the players $s = (s_1, \dots, s_I)$ induces an outcome of the game: a sequence of moves actually taken and a probability distribution over the terminal nodes of the game. Thus, for any profile of strategies $(s_1, \dots, s_I)$, we can deduce the payoffs received by each player. We might think, therefore, of specifying the game directly in terms of strategies and their associated payoffs. This second way to represent a game is known as the *normal* (or *strategic*) *form.* It is, in essence, a condensed version of the extensive form.

Player 2

Player 1		s_1	s_2	s_3	s_4
	H	−1, +1	−1, +1	+1, −1	+1, −1
	T	+1, −1	−1, +1	+1, −1	−1, +1

Figure 7.D.1
The normal [illegible] Matching P[illegible] Version B.

Definition 7.D.2: For a game with I players, the *normal form representation* Γ_N specifies for each player i a set of strategies S_i (with $s_i \in S_i$) and a payoff function $u_i(s_1, \ldots, s_I)$ giving the von Neumann–Morgenstern utility levels associated with the (possibly random) outcome arising from strategies $(s_1, \ldots, s_I)$. Formally, we write $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$.

In fact, when describing a game in its normal form, there is no need to keep track of the specific moves associated with each strategy. Instead, we can simply number the various possible strategies of a player, writing player i's strategy set as $S_i = \{s_{1i}, s_{2i}, \ldots\}$ and then referring to each strategy by its number.

A concrete example of a game in normal form is presented in Example 7.D.3 for Matching Pennies Version B.

Example 7.D.3: *The Normal Form of Matching Pennies Version B.* We have already described the strategy sets of the two players in Example 7.D.1. The payoff functions are

$$u_1(s_1, s_2) = \begin{cases} +1 \text{ if } (s_1, s_2) = (\text{H, strategies 3 or 4}) \text{ or } (\text{T, strategies 1 or 3}), \\ -1 \text{ if } (s_1, s_2) = (\text{H, strategies 1 or 2}) \text{ or } (\text{T, strategies 2 or 4}), \end{cases}$$

and $u_2(s_1, s_2) = -u_1(s_1, s_2)$. A convenient way to summarize this information is in the "game box" depicted in Figure 7.D.1. The different rows correspond to the strategies of player 1, and the columns to those of player 2. Within each cell, the payoffs of the two players are depicted as $(u_1(s_1, s_2), u_2(s_1, s_2))$. ■

Exercise 7.D.2: Depict the normal forms for Matching Pennies Version C and the standard version of Matching Pennies.

The idea behind using the normal form representation to study behavior in a game is that a player's decision problem can be thought of as one of choosing her strategy (her contingent plan of action) given the strategies that she thinks her rivals will be adopting. Because each player is faced with this problem, we can think of the players as simultaneously choosing their strategies from the sets $\{S_i\}$. It is as if the players each simultaneously write down their strategies on slips of paper and hand them to a referee, who then computes the outcome of the game from the players' submitted strategies.

From the previous discussion, it is clear that for any extensive form representation of a game, there is a unique normal form representation (more precisely, it is unique up to any renaming or renumbering of the strategies). The converse is not true, however. Many different extensive forms may be represented by the same normal form. For example, the normal form shown in Figure 7.D.1 represents not only the extensive form in Figure 7.C.1 but also the

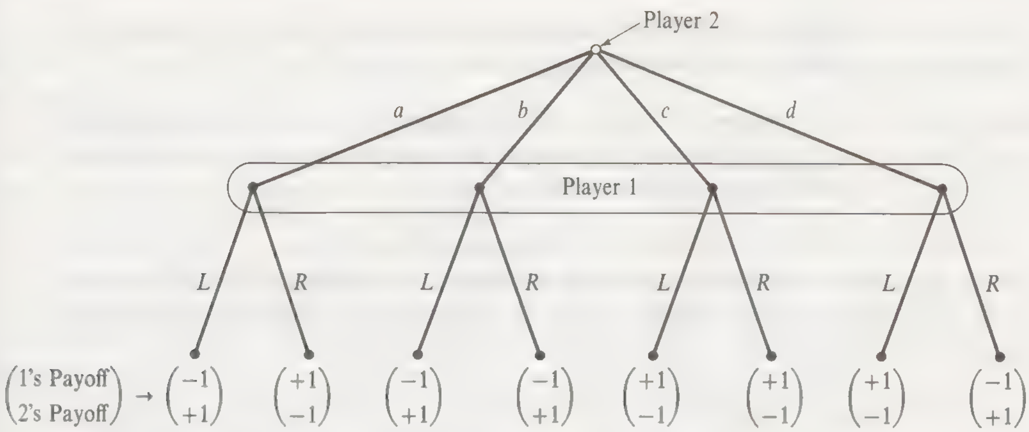


Figure 7.D.2 An extensive form whose normal form is that depicted in Figure 7.D.1.

extensive form in Figure 7.D.2. In the latter game, players move simultaneously, player 1 choosing between two strategies, L and R, and player 2 choosing among four strategies: a, b, c, and d. In terms of their representations in a game box, the only difference between the normal forms for these games lies in the "labels" given to the rows and columns.

Because the condensed representation of the game in the normal form generally omits some of the details present in the extensive form, we may wonder whether this omission is important or whether the normal form summarizes all of the strategically relevant information (as the last paragraph in regular type seems to suggest). The question can be put a little differently: Is the scenario in which players simultaneously write down their strategies and submit them to a referee really equivalent to their playing the game over time as described in the extensive form? This question is currently a subject of some controversy among game theorists. The debate centers on issues arising in dynamic games such as those studied in Chapter 9.

For the simultaneous-move games that we study in Chapter 8, in which all players choose their actions at the same time, the normal form captures *all* the strategically relevant information. In simultaneous-move games, a player's strategy is a simple noncontingent choice of an action. In this case, players' simultaneous choice of strategies in the normal form is clearly equivalent to their simultaneous choice of actions in the extensive form (captured there by having players not observing each other's choices).

Randomized Choices

Up to this point, we have assumed that players make their choices with certainty. However, there is no a priori reason to exclude the possibility that a player could randomize when faced with a choice. Indeed, we will see in Chapters 8 and 9 that in certain circumstances the possibility of randomization can play an important role in the analysis of games.

As stated in Definition 7.D.1, a deterministic strategy for player i, which we now call a *pure strategy*, specifies a deterministic choice $s_i(H)$ at each of her information sets $H \in \mathscr{H}_i$. Suppose that player i's (finite) set of pure strategies is S_i. One way for

the player to randomize is to choose randomly one element of this set. This kind of randomization gives rise to what is called a *mixed strategy*.

Definition 7.E.1: Given player i's (finite) pure strategy set S_i, a *mixed strategy* for player i, $\sigma_i: S_i \to [0, 1]$, assigns to each pure strategy $s_i \in S_i$ a probability $\sigma_i(s_i) \geq 0$ that it will be played, where $\sum_{s_i \in S_i} \sigma_i(s_i) = 1$.

Suppose that player i has M pure strategies in set $S_i = \{s_{1i}, \ldots, s_{Mi}\}$. Player i's set of possible mixed strategies can therefore be associated with the points of the following simplex (recall our use of a simplex to represent lotteries in Chapter 6):

$$\Delta(S_i) = \{(\sigma_{1i}, \ldots, \sigma_{Mi}) \in \mathbb{R}^M: \sigma_{mi} \geq 0 \text{ for all } m = 1, \ldots, M \text{ and } \sum_{m=1}^{M} \sigma_{mi} = 1\}.$$

This simplex is called the *mixed extension* of S_i. Note that a pure strategy can be viewed as a special case of a mixed strategy in which the probability distribution over the elements of S_i is degenerate.

When players randomize over their pure strategies, the induced outcome is itself random, leading to a probability distribution over the terminal nodes of the game. Since each player i's normal form payoff function $u_i(s)$ is of the von Neumann–Morgenstern type, player i's payoff given a profile of mixed strategies $\sigma = (\sigma_1, \ldots, \sigma_I)$ for the I players is her expected utility $E_\sigma[u_i(s)]$, the expectation being taken with respect to the probabilities induced by σ on pure strategy profiles $s = (s_1, \ldots, s_I)$. That is, letting $S = S_1 \times \cdots \times S_I$, player i's von Neumann–Morgenstern utility from mixed strategy profile σ is

$$\sum_{s \in S} [\sigma_1(s_1)\, \sigma_2(s_2) \cdots \sigma_I(s_I)]\, u_i(s),$$

which, with a slight abuse of notation, we denote by $u_i(\sigma)$. Note that because we assume that each player randomizes on her own, we take the realizations of players' randomizations to be independent of one another.[3]

The basic definition of the normal form representation need not be changed to accommodate the possibility that players might choose to play mixed strategies. We can simply consider the normal form game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ in which players' strategy sets are extended to include both pure and mixed strategies.

Note that we can equivalently think of a player forming her mixed strategy as follows: Player i has access to a private signal θ_i that is uniformly distributed on the interval $[0, 1]$ and is independent of other players' signals, and she forms her mixed strategy by making her plan of action contingent on the realization of the signal. That is, she specifies a pure strategy $s_i(\theta_i) \in S_i$ for each realization of θ_i. We shall return to this alternative interpretation of mixed strategies in Chapter 8.

If we use the extensive form description of a game, there is another way that player i could randomize. Rather than randomizing over the potentially very

3. In Chapter 8, however, we discuss the possibility that players' randomizations could be correlated.

set of pure strategies in S_i, she could randomize separately over the possible at each of her information sets $H \in \mathscr{H}_i$. This way of randomizing is called a *strategy*.

7.E.2: Given an extensive form game Γ_E, a *behavior strategy* for player specifies, for every information set $H \in \mathscr{H}_i$ and action $a \in C(H)$, a probability $H) \geq 0$, with $\sum_{a \in C(H)} \lambda_i(a, H) = 1$ for all $H \in \mathscr{H}_i$.

As might seem intuitive, for games of perfect recall (and we deal only with), the two types of randomization are equivalent. For any behavior strategy player i, there is a mixed strategy for that player that yields exactly the distribution over outcomes for any strategies, mixed or behavior, that might played by i's rivals, and vice versa [this result is due to Kuhn (1953); see 7.E.1]. Which form of randomized strategy we consider is therefore a analytical convenience; we typically use behavior strategies when analyzing form representation of a game and mixed strategies when analyzing normal form.

the way we introduce randomization is solely a matter of analytical we shall be a bit loose in our terminology and refer to all randomized as *mixed strategies*.

REFERENCES

R. (1976). Agreeing to disagree. *Annals of Statistics* **4**: 1236–39.

D. M., and R. Wilson. (1982). Sequential equilibrium. *Econometrica* **50**: 863–94.

H. W. (1953). Extensive games and the problem of information. In *Contributions to the Theory of Games*, vol. 2, edited by H. W. Kuhn and A. W. Tucker. Princeton, N.J.: Princeton University Press, 193–216.

P. (1981). An axiomatic characterization of common knowledge. *Econometrica* **49**: 219–22.

EXERCISES

Suppose that in the Meeting in New York game (Example 7.B.3), there are two possible where the two players can meet: Grand Central Station and the Empire State Building. an extensive form representation (game tree) for this game.

In a game where player i has N information sets indexed $n = 1, \ldots, N$ and M_n possible at information set n, how many strategies does player i have?

In text.

Consider the two-player game whose extensive form representation (excluding payoffs) in Figure 7.Ex.1.

(a) What are player 1's possible strategies? Player 2's?

(b) Show that for any behavior strategy that player 1 might play, there is a realization mixed strategy; that is, a mixed strategy that generates the same probability over the terminal nodes for *any* mixed strategy choice by player 2.

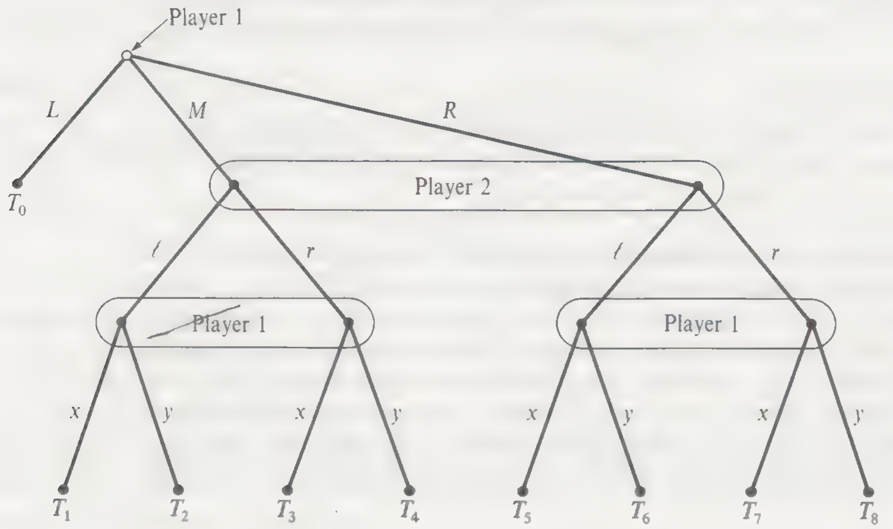


Figure 7.Ex.1

(c) Show that the converse is also true: For any mixed strategy that player 1 might play, there is a realization equivalent behavior strategy.

(d) Suppose that we change the game by merging the information sets at player 1's second round of moves (so that all four nodes are now in a single information set). Argue that the game is no longer one of perfect recall. Which of the two results in **(b)** and **(c)** still holds?

CHAPTER

8

Simultaneous-Move Games

8.A Introduction

We now turn to the central question of game theory: What should we expect to observe in a game played by rational players who are fully knowledgeable about the structure of the game and each others' rationality? In this chapter, we study *simultaneous-move* games, in which all players move only once and at the same time. Our motivation for beginning with these games is primarily pedagogic; they allow us to concentrate on the study of strategic interaction in the simplest possible setting and to defer until Chapter 9 some difficult issues that arise in more general, dynamic games.

In Section 8.B, we introduce the concepts of *dominant* and *dominated* strategies. These notions and their extension in the concept of *iterated dominance* provide a first and compelling restriction on the strategies rational players should choose to play. In Section 8.C, we extend these ideas by defining the notion of a *rationalizable strategy*. We argue that the implication of players' common knowledge of each others' rationality and of the structure of the game is precisely that they will play rationalizable strategies.

Unfortunately, in many games, the set of rationalizable strategies does not yield a very precise prediction of the play that will occur. In the remaining sections of the chapter, we therefore study solution concepts that yield more precise predictions by adding "equilibrium" requirements regarding players' behavior.

Section 8.D begins our study of equilibrium-based solution concepts by introducing the important and widely applied concept of *Nash equilibrium*. This concept adds to the assumption of common knowledge of players' rationality a requirement of *mutually correct expectations*. By doing so, it often greatly narrows the set of predicted outcomes of a game. We discuss in some detail the reasonableness of this requirement, as well as the conditions under which we can be assured that a Nash equilibrium exists.

In Sections 8.E and 8.F, we examine two extensions of the Nash equilibrium concept. In Section 8.E, we broaden the notion of a Nash equilibrium to cover situations with *incomplete information*, where each player's payoffs may, to some extent, be known only by the player. This yields the concept of *Bayesian Nash*

equilibrium. In Section 8.F, we explore the implications of players entertaining the possibility that, with some small but positive probability, their opponents might make a mistake in choosing their strategies. We define the notion of a (*normal form*) *trembling-hand perfect Nash equilibrium,* an extension of the Nash equilibrium concept that requires that equilibria be robust to the possibility of small mistakes.

Throughout the chapter, we study simultaneous-move games using their normal form representations (see Section 7.D). Thus, we use $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ when we consider only pure (nonrandom) strategy choices and $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ when we allow for the possibility of randomized choices by the players (see Section 7.E for a discussion of randomized choices). We often denote a profile of pure strategies for player i's opponents by $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_I)$, with a similar meaning applying to the profile of mixed strategies σ_{-i}. We then write $s = (s_i, s_{-i})$ and $\sigma = (\sigma_i, \sigma_{-i})$. We also let $S = S_1 \times \cdots \times S_I$ and $S_{-i} = S_1 \times \cdots \times S_{i-1} \times S_{i+1} \times \cdots \times S_I$.

8.B Dominant and Dominated Strategies

We begin our study of simultaneous-move games by considering the predictions that can be made based on a relatively simple means of comparing a player's possible strategies: that of *dominance.*

To keep matters as simple as possible, we initially ignore the possibility that players might randomize in their strategy choices. Hence, our focus is on games $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ whose strategy sets allow for only pure strategies.

Consider the game depicted in Figure 8.B.1, the famous *Prisoner's Dilemma.* The story behind this game is as follows: Two individuals are arrested for allegedly engaging in a serious crime and are held in separate cells. The district attorney (the DA) tries to extract a confession from each prisoner. Each is privately told that if he is the only one to confess, then he will be rewarded with a light sentence of 1 year while the recalcitrant prisoner will go to jail for 10 years. However, if he is the only one not to confess, then it is he who will serve the 10-year sentence. If both confess, they will both be shown some mercy: they will each get 5 years. Finally, if neither confesses, it will still be possible to convict both of a lesser crime that carries a sentence of 2 years. Each player wishes to minimize the time he spends in jail (or maximize the negative of this, the payoffs that are depicted in Figure 8.B.1).

What will the outcome of this game be? There is only one plausible answer: (confess, confess). To see why, note that playing "confess" is each player's best strategy *regardless of what the other player does.* This type of strategy is known as a *strictly dominant strategy.*

		Prisoner 2	
		Don't Confess	Confess
Prisoner 1	Don't Confess	$-2, -2$	$-10, -1$
	Confess	$-1, -10$	$-5, -5$

Figure 8.B.1
The Prisoner's Dilemma.

…on 8.B.1: A strategy $s_i \in S_i$ is a *strictly dominant strategy* for player i in game … $= [I, \{S_i\}, \{u_i(\cdot)\}]$ if for all $s_i' \neq s_i$, we have

$$u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$$

… all $s_{-i} \in S_{-i}$.

… words, a strategy s_i is a strictly dominant strategy for player i if it maximizes …ely player i's payoff for any strategy that player i's rivals might play. (The reason … modifier *strictly* in Definition 8.B.1 will be made clear in Definition 8.B.3.) …ayer has a strictly dominant strategy, as in the Prisoner's Dilemma, we should … him to play it.

… striking aspect of the (confess, confess) outcome in the Prisoner's Dilemma … although it is the one we expect to arise, it is not the best outcome for the … *jointly*; both players would prefer that neither of them confess. For this reason, … Prisoner's Dilemma is the paradigmatic example of self-interested, rational …avior *not* leading to a socially optimal result.

… way of viewing the outcome of the Prisoner's Dilemma is that, in seeking …imize his own payoff, each prisoner has a negative effect on his partner; by … away from the (don't confess, don't confess) outcome, a player reduces his … by 1 year but increases that of his partner by 8 (in Chapter 11, we shall see … as an example of an *externality*).

…ated Strategies

…gh it is compelling that players should play strictly dominant strategies if they … them, it is rare for such strategies to exist. Often, one strategy of player i's may … when his rivals play s_{-i} and another when they play some other strategies … of the standard Matching Pennies game in Chapter 7). Even so, we might … able to use the idea of dominance to eliminate some strategies as possible … In particular, we should expect that player i will not play *dominated* strategies, … which there is some alternative strategy that yields him a greater payoff …dless of what the other players do.

… 8.B.2: A strategy $s_i \in S_i$ is *strictly dominated* for player i in game … $S_i\}, \{u_i(\cdot)\}]$ if there exists another strategy $s_i' \in S_i$ such that for all … $\in S$…

$$u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}).$$

…this case, we say that strategy s_i' *strictly dominates* strategy s_i.

… this definition, we can restate our definition of a strictly dominant strategy …n 8.B.1) as follows: Strategy s_i is a strictly dominant strategy for player i in … $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if it strictly dominates every other strategy in S_i.

…ple 8.B.1: Consider the game shown in Figure 8.B.2. There is no strictly … strategy, but strategy D for player 1 is strictly dominated by strategy M … also by strategy U). ■

…ition 8.D.3 presents a related, weaker notion of a dominated strategy that … some importance.

Player 2

Player 1		L	R
	U	1, −1	−1, 1
Player 1	M	−1, 1	1, −1
	D	−2, 5	−3, 2

Figure 8.B.2
Strategy D is
dominated.

Player 2

Player 1		L	R
	U	5, 1	4, 0
Player 1	M	6, 0	3, 1
	D	6, 4	4, 4

Figure 8.B.3
Strategies U
are weakly

Definition 8.B.3: A strategy $s_i \in S_i$ is *weakly dominated* in game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if there exists another strategy $s_i' \in S_i$ such that for all $s_{-i} \in S_{-i}$,

$$u_i(s_i', s_{-i}) \geq u_i(s_i, s_{-i}),$$

with strict inequality for *some* s_{-i}. In this case, we say that strategy s_i' *weakly dominates* strategy s_i. A strategy is a *weakly dominant strategy* for player i in game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if it weakly dominates every other strategy in S_i.

Thus, a strategy is weakly dominated if another strategy does at least as well for all s_{-i} and strictly better for some s_{-i}.

Example 8.B.2: Figure 8.B.3 depicts a game in which player 1 has two weakly dominated strategies, U and M. Both are weakly dominated by strategy D. ■

Unlike a strictly dominated strategy, a strategy that is only weakly dominated cannot be ruled out based solely on principles of rationality. For any alternative strategy that player i might pick, there is at least one profile of strategies for his rivals for which the weakly dominated strategy does as well. In Figure 8.B.3, for example, player 1 could rationally pick M if he was *absolutely sure* that player 2 would play L. Yet, if the probability of player 2 choosing strategy R was perceived by player 1 as positive (no matter how small), then M would not be a rational choice for player 1. *Caution* might therefore rule out M. More generally, weakly dominated strategies could be dismissed if players always believed that there was at least some positive probability that any strategies of their rivals could be chosen. We do not pursue this idea here, although we return to it in Section 8.F. For now, we continue to allow a player to entertain any conjecture about what an opponent might play, even a perfectly certain one.

Iterated Deletion of Strictly Dominated Strategies

As we have noted, it is unusual for elimination of strictly dominated strategies to lead to a unique prediction for a game (e.g., recall the game in Figure 8.B.2). However, the logic of eliminating strictly dominated strategies can be pushed further, as demonstrated in Example 8.B.3.

Example 8.B.3: In Figure 8.B.4, we depict a modification of the Prisoner's Dilemma, which we call the *DA's Brother.*

The story (a somewhat far-fetched one!) is now as follows: One of the prisoners, prisoner 1, is the DA's brother. The DA has some discretion in the fervor with which

		Prisoner 2	
		Don't Confess	Confess
Prisoner 1	Don't Confess	0, −2	−10, −1
	Confess	−1, −10	−5, −5

Figure 8.B.4
The DA's Brother.

[illegible] ecutes and, in particular, can allow prisoner 1 to go free if neither of the [illegible] confesses. With this change, if prisoner 2 confesses, then prisoner 1 should [illegible]ess; but "don't confess" has become prisoner 1's best strategy if prisoner 2 [illegible] n't confess." Thus, we are unable to rule out either of prisoner 1's strategies [illegible] dominated, and so elimination of strictly dominated (or, for that matter, [illegible] dominated) strategies does not lead to a unique prediction.

[illegible]ever, we can still derive a unique prediction in this game if we push the logic [illegible]ating strictly dominated strategies further. Note that "don't confess" is still [illegible] dominated for prisoner 2. Furthermore, once prisoner 1 eliminates "don't [illegible]" as a possible action by prisoner 2, "confess" is prisoner 1's unambiguously [illegible] action; that is, it is his strictly dominant strategy once the strictly dominated [illegible] of prisoner 2 has been deleted. Thus, the unique predicted outcome in the [illegible]'s Brother game should still be (confess, confess). ■

[illegible] the way players' common knowledge of each other's payoffs and rationality [illegible] to solve the game in Example 8.B.3. Elimination of strictly dominated [illegible]es requires only that each player be rational. What we have just done, [illegible], requires not only that prisoner 2 be rational but also that prisoner 1 *know* [illegible]oner 2 is rational. Put somewhat differently, a player need not know anything [illegible] opponents' payoffs or be sure of their rationality to eliminate a strictly [illegible]ed strategy from consideration as his own strategy choice; but for the player [illegible]ate one of his strategies from consideration because it is dominated if his [illegible]nts never play *their* dominated strategies *does* require this knowledge.

[illegible] general matter, if we are willing to assume that all players are rational *and* [illegible] fact and the players' payoffs are common knowledge (so everybody knows [illegible] everybody knows that ... everybody is rational), then we do not need to stop [illegible] two iterations. We can eliminate not only strictly dominated strategies and [illegible] that are strictly dominated after the first deletion of strategies but also [illegible] that are strictly dominated after this *next* deletion of strategies, and so on. [illegible] with each elimination of strategies, it becomes possible for additional [illegible] to become dominated because the fewer strategies that a player's opponents [illegible]ay, the more likely that a particular strategy of his is dominated. However, [illegible]itional iteration requires that players' knowledge of each others' rationality [illegible] level deeper. A player must now know not only that his rivals are rational [illegible] also that they know that he is, and so on.

[illegible] feature of the process of iteratively eliminating strictly dominated strategies [illegible] the order of deletion does not affect the set of strategies that remain in the [illegible] Exercise 8.B.4), That is, if at any given point several strategies (of one or

several players) are strictly dominated, then we can eliminate them all at once or in any sequence without changing the set of strategies that we ultimately end up with. This is fortunate, since we would worry if our prediction depended on the arbitrarily chosen order of deletion.

Exercise 8.B.5 presents an interesting example of a game for which the iterated removal of strictly dominated strategies yields a unique prediction: the *Cournot duopoly game* (which we will discuss in detail in Chapter 12).

The iterated deletion of *weakly* dominated strategies is harder to justify. As we have already indicated, the argument for deletion of a weakly dominated strategy for player i is that he contemplates the possibility that every strategy combination of his rivals occurs with positive probability. However, this hypothesis clashes with the logic of iterated deletion, which assumes, precisely, that eliminated strategies are not expected to occur. This inconsistency leads the iterative elimination of weakly dominated strategies to have the undesirable feature that it *can* depend on the order of deletion. The game in Figure 8.B.3 provides an example. If we first eliminate strategy U, we next eliminate strategy L, and we can then eliminate strategy M; (D, R) is therefore our prediction. If, instead, we eliminate strategy M first, we next eliminate strategy R, and we can then eliminate strategy U; now (D, L) is our prediction.

Allowing for Mixed Strategies

When we recognize that players may randomize over their pure strategies, the basic definitions of strictly dominated and dominant strategies can be generalized in a straightforward way.

Definition 8.B.4: A strategy $\sigma_i \in \Delta(S_i)$ is *strictly dominated* for player i in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ if there exists another strategy $\sigma_i' \in \Delta(S_i)$ such that for all $\sigma_{-i} \in \prod_{j \neq i} \Delta(S_j)$,

$$u_i(\sigma_i', \sigma_{-i}) > u_i(\sigma_i, \sigma_{-i}).$$

In this case, we say that strategy σ_i' *strictly dominates* strategy σ_i. A strategy σ_i is a *strictly dominant strategy* for player i in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ if it strictly dominates every other strategy in $\Delta(S_i)$.

Using this definition and the structure of mixed strategies, we can say a bit more about the set of strictly dominated strategies in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$.

Note first that when we test whether a strategy σ_i is strictly dominated by strategy σ_i' for player i, we need only consider these two strategies' payoffs against the *pure* strategies of i's opponents. That is,

$$u_i(\sigma_i', \sigma_{-i}) > u_i(\sigma_i, \sigma_{-i}) \quad \text{for all } \sigma_{-i}$$

if and only if

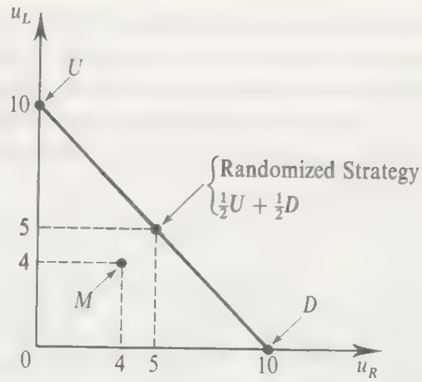
$$u_i(\sigma_i', s_{-i}) > u_i(\sigma_i, s_{-i}) \quad \text{for all } s_{-i}.$$

This follows because we can write

$$u_i(\sigma_i', \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i}) = \sum_{s_{-i} \in S_{-i}} \left[\prod_{k \neq i} \sigma_k(s_k) \right] [u_i(\sigma_i', s_{-i}) - u_i(\sigma_i, s_{-i})].$$

		Player 2	
		L	R
	U	10, 1	0, 4
Player 1	M	4, 2	4, 3
	D	0, 5	10, 2

(a)



(b)

Figure 8.B.5
Domination of a pure strategy by a randomized strategy.

This expression is positive for all σ_{-i} if and only if $[u_i(\sigma_i', s_{-i}) - u_i(\sigma_i, s_{-i})]$ is positive for all s_{-i}. One implication of this point is presented in Proposition 8.B.1.

Proposition 8.B.1: Player i's pure strategy $s_i \in S_i$ is strictly dominated in game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if and only if there exists another strategy $\sigma_i' \in \Delta(S_i)$ such that

$$u_i(\sigma_i', s_{-i}) > u_i(s_i, s_{-i})$$

for all $s_{-i} \in S_{-i}$.

Proposition 8.B.1 tells us that to test whether a pure strategy s_i is dominated when randomized play is possible, the test given in Definition 8.B.2 need only be augmented by checking whether any of player i's mixed strategies does better than s_i against every possible profile of pure strategies by i's rivals.

In fact, this extra requirement can eliminate additional pure strategies because a pure strategy s_i may be dominated only by a randomized combination of other pure strategies; that is, to dominate a strategy, even a pure one, it may be necessary to consider alternative strategies that involve randomization. To see this, consider the two-player game depicted in Figure 8.B.5(a). Player 1 has three strategies: U, M, and D. We can see that U is an excellent strategy when player 2 plays L but a poor one against R and that D is excellent against R and poor against L. Strategy M, on the other hand, is a good but not great strategy against both L and R. None of these three pure strategies is strictly dominated by any of the others. But if we allow player 1 to randomize, then playing U and D each with probability $\frac{1}{2}$ yields player 1 an expected payoff of 5 regardless of player 2's strategy, strictly dominating M (remember, payoffs are in terms of von Neumann Morgenstern utilities). This is shown in Figure 8.B.5(b), where player 1's expected payoffs from playing U, D, M, and the randomized strategy $\frac{1}{2}U + \frac{1}{2}D$ are plotted as points in $\mathbb{R}^2$ (the two dimensions correspond to a strategy's expected payoff for player 1 when player 2 plays R, denoted by u_R, and L, denoted by u_L). In the figure, the payoff vectors achievable by randomizing over U and D, and that from the randomized strategy $\frac{1}{2}U + \frac{1}{2}D$ in particular, lie on the line connecting points (0, 10) and (10, 0). As can be seen, the payoffs from $\frac{1}{2}U + \frac{1}{2}D$ strictly dominate those from strategy M.

Once we have determined the set of undominated pure strategies for player i, we need to consider which mixed strategies are undominated. We can immediately eliminate any mixed strategy that uses a dominated pure strategy; if pure strategy s_i is strictly dominated for player i, then so is every mixed strategy that assigns a positive probability to this strategy.

Exercise 8.B.6: Prove that if pure strategy s_i is a strictly dominated strategy in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, then so is any strategy that plays s_i with positive probability.

But these are not the only mixed strategies that may be dominated. A mixed strategy that randomizes over undominated pure strategies may itself be dominated. For example, if strategy M in Figure 8.B.5(a) instead gave player 1 a payoff of 6 for either strategy chosen by player 2, then although neither strategy U nor strategy D would be strictly dominated, the randomized strategy $\frac{1}{2}U + \frac{1}{2}D$ would be strictly dominated by strategy M [look where the point (6, 6) would lie in Figure 8.B.5(b)].

In summary, to find the set of strictly dominated strategies for player i in $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, we can first eliminate those pure strategies that are strictly dominated by applying the test in Proposition 8.B.1. Call player i's set of undominated pure strategies $S_i^u \subset S_i$. Next, eliminate any mixed strategies in set $\Delta(S_i^u)$ that are dominated. Player i's set of undominated strategies (pure and mixed) is exactly the remaining strategies in set $\Delta(S_i^u)$.

As when we considered only pure strategies, we can push the logic of removal of strictly dominated strategies in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ further through iterative elimination. The preceding discussion implies that this iterative procedure can be accomplished with the following two-stage procedure: First iteratively eliminate dominated pure strategies using the test in Proposition 8.B.1, applied at each stage using the remaining set of pure strategies. Call the remaining sets of pure strategies $\{\bar{S}_1^u, \ldots, \bar{S}_I^u\}$. Then, eliminate any mixed strategies in sets $\{\Delta(\bar{S}_1^u), \ldots, \Delta(\bar{S}_I^u)\}$ that are dominated.

8.C Rationalizable Strategies

In Section 8.B, we eliminated strictly dominated strategies based on the argument that a rational player would never choose such a strategy regardless of the strategies that he anticipates his rivals will play. We then used players' common knowledge of each others' rationality and the structure of the game to justify iterative removal of strictly dominated strategies.

In general, however, players' common knowledge of each others' rationality and the game's structure allows us to eliminate more than just those strategies that are iteratively strictly dominated. Here, we develop this point, leading to the concept of a *rationalizable strategy*. The set of rationalizable strategies consists precisely of those strategies that may be played in a game where the structure of the game and the players' rationality are common knowledge among the players. Throughout this section, we focus on games of the form $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ (mixed strategies are permitted).

We begin with Definition 8.C.1.

Definition 8.C.1: In game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, strategy σ_i is a *best response* for player i to his rivals' strategies σ_{-i} if

$$u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma_i', \sigma_{-i})$$

for all $\sigma_i' \in \Delta(S_i)$. Strategy σ_i is *never a best response* if there is no σ_{-i} for which σ_i is a best response.

Strategy σ_i is a best response to σ_{-i} if it is an optimal choice when player i conjectures that his opponents will play σ_{-i}. Player i's strategy σ_i is never a best response if there is no belief that player i may hold about his opponents' strategy

beliefs σ_{-i} that justifies choosing strategy σ_i.[1] Clearly, a player should not play a strategy that is never a best response.

Note that a strategy that is strictly dominated is never a best response. However, as a general matter, a strategy might never be a best response even though it is not strictly dominated (we say more about this relation at the end of this section in small type). Thus, eliminating strategies that are never a best response must eliminate at least as many strategies as eliminating just strictly dominated strategies and may eliminate more.

Moreover, as in the case of strictly dominated strategies, common knowledge of rationality and the game's structure implies that we can iterate the deletion of strategies that are never a best response. In particular, a rational player should not play a strategy that is never a best response once he eliminates the possibility that any of his rivals might play a strategy that is never a best response for them, and so on.

Equally important, the strategies that remain after this iterative deletion are the strategies that a rational player can *justify*, or *rationalize*, affirmatively with some reasonable conjecture about the choices of his rivals; that is, with a conjecture that does not assume that any player will play a strategy that is never a best response or one that is only a best response to a conjecture that someone else will play such a strategy, and so on. (Example 8.C.1 provides an illustration of this point.) As a result, the set of strategies surviving this iterative deletion process can be said to be precisely the set of strategies that can be played by rational players in a game in which the players' rationality and the structure of the game are common knowledge. They are known as *rationalizable strategies* [a concept developed independently by Bernheim (1984) and Pearce (1984)].

Definition 8.C.2: In game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, the strategies in $\Delta(S_i)$ that survive the iterated removal of strategies that are never a best response are known as *player i's rationalizable strategies.*

Note that the set of rationalizable strategies can be no larger than the set of strategies surviving iterative removal of strictly dominated strategies because, at each stage of the iterative process in Definition 8.C.2, all strategies that are strictly dominated at that stage are eliminated. As in the case of iterated deletion of strictly dominated strategies, the order of removal of strategies that are never a best response can be shown not to affect the set of strategies that remain in the end (see Exercise 8.C.2).

1. We speak here as if a player's conjecture is necessarily deterministic in the sense that the player believes it is certain that his rivals will play a particular profile of mixed strategies σ_{-i}. One might wonder about conjectures that are probabilistic, that is, that take the form of a nondegenerate probability distribution over possible profiles of mixed strategy choices by his rivals. In fact, a strategy σ_i is an optimal choice for player i given some probabilistic conjecture (that treats his opponents' choices as independent random variables) only if it is an optimal choice given some deterministic conjecture. The reason is that if σ_i is an optimal choice given some probabilistic conjecture, then it must be a best response to the profile of mixed strategies σ_{-i} that plays each possible pure strategy profile $s_{-i} \in S_{-i}$ with exactly the compound probability implied by the probabilistic conjecture.

Player 2

Player 1	b_1	b_2	b_3	b_4
a_1	0, 7	2, 5	7, 0	0, 1
a_2	5, 2	3, 3	5, 2	0, 1
a_3	7, 0	2, 5	0, 7	0, 1
a_4	0, 0	0, −2	0, 0	10, −1

Figure 8.C.1
$\{a_1, a_2,$ …
are ra…
strateg…
$\{b_1, b_2,$ …
ration…
strategie…

Example 8.C.1: Consider the game depicted in Figure 8.C.1, which is taken from Bernheim (1984). What is the set of rationalizable pure strategies for the two players? In the first round of deletion, we can eliminate strategy b_4, which is never a best response because it is strictly dominated by a strategy that plays strategies b_1 and b_3 each with probability $\frac{1}{2}$. Once strategy b_4 is eliminated, strategy a_4 can be eliminated because it is strictly dominated by a_2 once b_4 is deleted. At this point, no further strategies can be ruled out: a_1 is a best response to b_3, a_2 is a best response to b_2, and a_3 is a best response to b_1. Similarly, you can check that b_1, b_2, and b_3 are each best responses to one of a_1, a_2, and a_3. Thus, the set of rationalizable pure strategies for player 1 is $\{a_1, a_2, a_3\}$, and the set $\{b_1, b_2, b_3\}$ is rationalizable for player 2.

Note that for each of these rationalizable strategies, a player can construct a *chain of justification* for his choice that never relies on any player believing that another player will play a strategy that is never a best response.[2] For example, in the game in Figure 8.C.1, player 1 can justify choosing a_2 by the belief that player 2 will play b_2, which player 1 can justify to himself by believing that player 2 will think that he is going to play a_2, which is reasonable if player 1 believes that player 2 is thinking that he, player 1, thinks player 2 will play b_2, and so on. Thus, player 1 can construct an (infinite) chain of justification for playing strategy a_2, $(a_2, b_2, a_2, b_2, \ldots)$, where each element is justified using the next element in the sequence.

Similarly, player 1 can rationalize playing strategy a_1 with the chain of justification $(a_1, b_3, a_3, b_1, a_1, b_3, a_3, b_1, a_1, \ldots)$. Here player 1 justifies playing a_1 by believing that player 2 will play b_3. He justifies the belief that player 2 will play b_3 by thinking that player 2 believes that he, player 1, will play a_3. He justifies this belief by thinking that player 2 thinks that he, player 1, believes that player 2 will play b_1. And so on.

Suppose, however, that player 1 tried to justify a_4. He could do so only by a belief that player 2 would play b_4, but there is *no* belief that player 2 could have that would justify b_4. Hence, player 1 cannot justify playing the nonrationalizable strategy a_4. ■

2. In fact, this chain-of-justification approach to the set of rationalizable strategies is used in the original definition of the concept [for a formal treatment, consult Bernheim (1984) and Pearce (1984)].

It can be shown that under fairly weak conditions a player always has at least one rationalizable strategy.[3] Unfortunately, players may have many rationalizable strategies, as in Example 8.C.1. If we want to narrow our predictions further, we need to make additional assumptions beyond common knowledge of rationality. The solution concepts studied in the remainder of this chapter do so by imposing "equilibrium" requirements on players' strategy choices.

We have said that the set of rationalizable strategies is no larger than the set remaining after iterative deletion of strictly dominated strategies. It turns out, however, that for the case of two-player games ($I = 2$), these two sets are identical because in two-player games a (mixed) strategy σ_i is a best response to some strategy choice of a player's rival whenever σ_i is not strictly dominated.

To see that this is plausible, reconsider the game in Figure 8.B.5 (Exercise 8.C.3 asks you for a general proof). Suppose that the payoffs from strategy M are altered so that M is not strictly dominated. Then, as depicted in Figure 8.C.2, the payoffs from M lie somewhere above

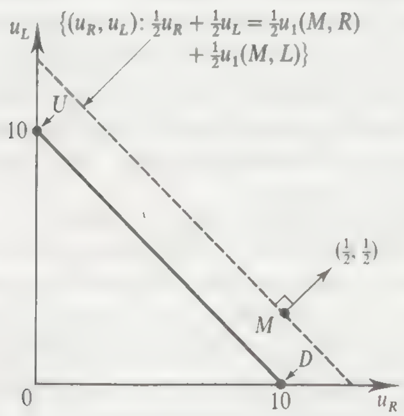


Figure 8.C.2
In a two-player game, a strategy is a best response if it is not strictly dominated.

the line connecting the points for strategies U and D. Is M a best response here? Yes. To see this, note that if player 2 plays strategy R with probability $\sigma_2(R)$, then player 1's expected payoff from choosing a strategy with payoffs (u_R, u_L) is $\sigma_2(R)u_R + (1 - \sigma_2(R))u_L$. Points yielding the same expected payoff as strategy M therefore lie on a hyperplane with normal vector $(1 - \sigma_2(R), \sigma_2(R))$. As can be seen, strategy M is a best response to $\sigma_2(R) = \frac{1}{2}$; it yields an expected payoff strictly larger than any expected payoff achievable by playing strategies U and/or D.

With more than two players, however, there can be strategies that are never a best response and yet are not strictly dominated. The reason can be traced to the fact that players' randomizations are independent. If the randomizations by i's rivals can be correlated (we discuss how this might happen at the end of Sections 8.D and 8.E), the equivalence reemerges. Exercise 8.C.4 illustrates these points.

---

3. This will be true, for example, whenever a Nash equilibrium (introduced in Section 8.D) exists.

8.D Nash Equilibrium

In this section, we present and discuss the most widely used solution concept in applications of game theory to economics, that of *Nash equilibrium* [due to Nash (1951)]. Throughout the rest of the book, we rely on it extensively.

For ease of exposition, we initially ignore the possibility that players might randomize over their pure strategies, restricting our attention to game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$. Mixed strategies are introduced later in the section.

We begin with Definition 8.D.1.

Definition 8.D.1: A strategy profile $s = (s_1, \ldots, s_I)$ constitutes a *Nash equilibrium* of game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if for every $i = 1, \ldots, I$,

$$u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$$

for all $s_i' \in S_i$.

In a Nash equilibrium, each player's strategy choice is a best response (see Definition 8.C.1) to the strategies *actually played* by his rivals. The italicized words distinguish the concept of Nash equilibrium from the concept of rationalizability studied in Section 8.C. Rationalizability, which captures the implications of the players' common knowledge of each others' rationality and the structure of the game, requires only that a player's strategy be a best response to some reasonable conjecture about what his rivals will be playing, where *reasonable* means that the conjectured play of his rivals can also be so justified. Nash equilibrium adds to this the requirement that players be *correct* in their conjectures.

Examples 8.D.1 and 8.D.2 illustrate the use of the concept.

Example 8.D.1: Consider the two-player simultaneous-move game shown in Figure 8.D.1. We can see that (M, m) is a Nash equilibrium. If player 1 chooses M, then the best response of player 2 is to choose m; the reverse is true for player 2. Moreover, (M, m) is the only combination of (pure) strategies that is a Nash equilibrium. For example, strategy profile (U, r) cannot be a Nash equilibrium because player 1 would prefer to deviate to strategy D given that player 2 is playing r. (Check the other possibilities for yourself.) ■

Example 8.D.2: *Nash Equilibrium in the Game of Figure 8.C.1.* In this game, the unique Nash equilibrium profile of (pure) strategies is (a_2, b_2). Player 1's best response to b_2 is a_2, and player 2's best response to a_2 is b_2, so (a_2, b_2) is a Nash equilibrium.

Player 2

		ℓ	m	r
	U	5, 3	0, 4	3, 5
Player 1	M	4, 0	5, 5	4, 0
	D	3, 5	0, 4	5, 3

Figure 8.D.1
A Nash equili

		Mr. Schelling	
		Empire State	Grand Central
Mr. Thomas	Empire State	100, 100	0, 0
	Grand Central	0, 0	100, 100

Figure 8.D.2
Nash equilibria in the Meeting in New York game.

other strategy profile, one of the players has an incentive to deviate. [In fact, is the unique Nash equilibrium even when randomization is permitted; Exercise 8.D.1.]

example illustrates a general relationship between the concept of Nash brium and that of rationalizable strategies: *Every strategy that is part of a uilibrium profile is rationalizable* because each player's strategy in a Nash um can be justified by the Nash equilibrium strategies of the other players. a general matter, the Nash equilibrium concept offers at least as sharp a n as does the rationalizability concept. In fact, it often offers a *much* sharper n. In the game of Figure 8.C.1, for example, the rationalizable strategies a_1, and b_3 are eliminated as predictions because they cannot be sustained when ers' beliefs about each other's play are required to be correct. ■

e previous two examples, the Nash equilibrium concept yields a unique n. However, this is not always the case. Consider the Meeting in New York

8.D.3: *Nash Equilibria in the Meeting in New York Game.* Figure 8.D.2 simple version of the Meeting in New York game. Mr. Thomas and Mr. ng each have two choices: They can meet either at noon at the top of the State Building or at noon at the clock in Grand Central Station. There are h equilibria (ignoring the possibility of randomization): (Empire State, State) and (Grand Central, Grand Central). ■

ple 8.D.3 emphasizes how strongly the Nash equilibrium concept uses the n of mutually correct expectations. The theory of Nash equilibrium is silent equilibrium we should expect to see when there are many. Yet, the players ssumed to correctly forecast which one it will be.

pact restatement of the definition of a Nash equilibrium can be obtained the introduction of the concept of a player's *best-response correspondence*. ally, we say that player i's best-response correspondence $b_i: S_{-i} \rightarrow S_i$ in game $\Gamma_N = [I, \{S_i\} \{u_i(\cdot)\}]$, is the correspondence that assigns to each $s_{-i} \in S_{-i}$

$$b_i(s_{-i}) = \{s_i \in S_i: u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \text{ for all } s_i' \in S_i\}.$$

notion, we can restate the definition of a Nash equilibrium as follows: The profile $(s_1, \ldots, s_I)$ is a Nash equilibrium of game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if ly if $s_i \in b_i(s_{-i})$ for $i = 1, \ldots, I$.

Discussion of the Concept of Nash Equilibrium

Why might it be reasonable to expect players' conjectures about each other's play to be correct? Or, in sharper terms, why should we concern ourselves with the concept of Nash equilibrium?

A number of arguments have been put forward for the Nash equilibrium concept and you will undoubtedly react to them with varying degrees of satisfaction. Moreover, one argument might seem compelling in one application but not at all convincing in another. Until very recently, all these arguments have been informal, as will be our discussion. The issue is one of the more important open questions in game theory, particularly given the Nash equilibrium concept's widespread use in applied problems, and it is currently getting some formal attention.

(i) *Nash equilibrium as a consequence of rational inference.* It is sometimes argued that because each player can think through the strategic considerations faced by his opponents, rationality alone implies that players must be able to correctly forecast what their rivals will play. Although this argument may seem appealing, it is faulty. As we saw in Section 8.C, the implication of common knowledge of the players' rationality (and of the game's structure) is precisely that each player must play a rationalizable strategy. Rationality need not lead players' forecasts to be correct.

(ii) *Nash equilibrium as a necessary condition if there is a unique predicted outcome to a game.* A more satisfying version of the previous idea argues that if there is a unique predicted outcome for a game, then rational players will understand this. Therefore, for no player to wish to deviate, this predicted outcome must be a Nash equilibrium. Put somewhat differently [as in Kreps (1990)], if players think and share the belief that there is an *obvious* (in particular, a unique) way to play a game, then it must be a Nash equilibrium.

Of course, this argument is only relevant if there is a unique prediction for how players will play a game. The discussion of rationalizability in Section 8.C, however, shows that common knowledge of rationality alone does not imply this. Therefore, this argument is really useful only in conjunction with some reason why a particular profile of strategies might be the obvious way to play a particular game. The other arguments for Nash equilibrium that we discuss can be viewed as combining this argument with a reason why there might be an "obvious" way to play a game.

(iii) *Focal points.* It sometimes happens that certain outcomes are what Schelling (1960) calls *focal.* For example, take the Meeting in New York game depicted in Figure 8.D.2, and suppose that restaurants in the Grand Central area are so much better then those around the Empire State Building that the payoffs to meeting at Grand Central are (1000, 1000) rather than (100, 100). Suddenly, going to Grand Central seems like the obvious thing to do. Focal outcomes could also be culturally determined. As Schelling pointed out in his original discussion, two people who do not live in New York will tend to find meeting at the top of the Empire State building (a famous tourist site) to be focal, whereas two native New Yorkers will find Grand

al Station (the central railroad station) a more compelling choice. In both ples, one of the outcomes has a natural appeal. The implication of argument that this kind of appeal can lead an outcome to be the clear prediction in a only if the outcome is a Nash equilibrium.

Nash equilibrium as a self-enforcing agreement. Another argument for Nash rium comes from imagining that the players can engage in nonbinding communication prior to playing the game. If players agree to an outcome to be played, aturally becomes the obvious candidate for play. However, because players bind themselves to their agreed-upon strategies, any agreement that the reach must be self-enforcing if it is to be meaningful. Hence, any meaningful ment must involve the play of a Nash equilibrium strategy profile. Of course, hough players have reached an agreement to play a Nash equilibrium, they still deviate from it if they expect others to do so. In essence, this justification that once the players have agreed to a choice of strategies, this agreement mes focal.

(v) *Nash equilibrium as a stable social convention.* A particular way to play a might arise over time if the game is played repeatedly and some stable social ntion emerges. If it does, it may be "obvious" to all players that the convention be maintained. The convention, so to speak, becomes focal.

good example is the game played by New Yorkers every day: Walking in own Manhattan. Every day, people who walk to work need to decide which the sidewalk they will walk on. Over time, the stable social convention is that ne walks on the right side, a convention that is enforced by the fact that any ual who unilaterally deviates from it is sure to be severely trampled. Of course, given day, it is *possible* that an individual might decide to walk on the left njecturing that everyone else suddenly expects the convention to change. heless, the prediction that we will remain at the Nash equilibrium "everyone s on the right" seems reasonable in this case. Note that if an outcome is to a stable social convention, it must be a Nash equilibrium. If it were not, then iduals would deviate from it as soon as it began to emerge.

e notion of an equilibrium as a rest point for some dynamic adjustment process es the use and the traditional appeal of equilibrium notions in economics. In se, the stable social convention justification of Nash equilibrium is closest to tradition of economic theory.

---

formally model the emergence of stable social conventions is not easy. One difficulty is the repeated one-day game may itself be viewed as a larger dynamic game. Thus, when consider rational players choosing their strategies in this overall game, we are merely led to our original conundrum: Why should we expect a Nash equilibrium in this larger ? One response to this difficulty currently getting some formal attention imagines that follow simple rules of thumb concerning their opponents' likely play in situations where is repeated (note that this implies a certain withdrawal from the assumption of complete). For example, a player could conjecture that whatever his opponents did yesterday be repeated today. If so, then each day players will play a best response to yesterday's . If a combination of strategies arises that is a stationary point of this process (i.e., the

		Player 2	
		Heads	Tails
Player 1	Heads	$-1, +1$	$+1, -1$
	Tails	$+1, -1$	$-1, +1$

Figure 8.D.3 Matching Pennies.

play today is the same as it was yesterday), it must be a Nash equilibrium. However, it is less clear that from any initial position, the process will converge to a stationary outcome; convergence turns out to depend on the game.[4]

Mixed Strategy Nash Equilibria

It is straightforward to extend the definition of Nash equilibrium to games in which we allow the players to randomize over their pure strategies.

Definition 8.D.2: A mixed strategy profile $\sigma = (\sigma_1, \ldots, \sigma_I)$ constitutes a *Nash equilibrium* of game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ if for every $i = 1, \ldots, I$,

$$u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma_i', \sigma_{-i})$$

for all $\sigma_i' \in \Delta(S_i)$.

Example 8.D.4: As a very simple example, consider the standard version of Matching Pennies depicted in Figure 8.D.3. This is a game with no pure strategy equilibrium. On the other hand, it is fairly intuitive that there is a mixed strategy equilibrium in which each player chooses H or T with equal probability. When a player randomizes in this way, it makes his rival indifferent between playing heads or tails, and so his rival is also willing to randomize between heads and tails with equal probability. ■

It is not an accident that a player who is randomizing in a Nash equilibrium of Matching Pennies is indifferent between playing heads and tails. As Proposition 8.D.1 confirms, this indifference among strategies played with positive probability is a general feature of mixed strategy equilibria.

Proposition 8.D.1: Let $S_i^+ \subset S_i$ denote the set of pure strategies that player i plays with positive probability in mixed strategy profile $\sigma = (\sigma_1, \ldots, \sigma_I)$. Strategy profile σ is a Nash equilibrium in game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ if and only if for all $i = 1, \ldots, I$,

(i) $u_i(s_i, \sigma_{-i}) = u_i(s_i', \sigma_{-i})$ for all $s_i, s_i' \in S_i^+$;
(ii) $u_i(s_i, \sigma_{-i}) \geq u_i(s_i', \sigma_{-i})$ for all $s_i \in S_i^+$ and all $s_i' \notin S_i^+$.

Proof: For necessity, note that if either of conditions (i) or (ii) does not hold for some player i, then there are strategies $s_i \in S_i^+$ and $s_i' \in S_i$ such that $u_i(s_i', \sigma_{-i}) > u_i(s_i, \sigma_{-i})$. If so, player i can strictly increase his payoff by playing strategy s_i' whenever he would have played strategy s_i.

---

4. This approach actually dates to Cournot's (1838) myopic adjustment procedure. A recent example can be found in Milgrom and Roberts (1990). Interestingly, this work explains the "ultrarational" Nash outcome by *relaxing* the assumption of rationality. It also can be used to try to identify the likelihood of various Nash equilibria arising when multiple Nash equilibria exist.

For sufficiency, suppose that conditions (i) and (ii) hold but that σ is not a Nash equilibrium. Then there is some player i who has a strategy σ_i' with $u_i(\sigma_i', \sigma_{-i}) > u_i(\sigma_i, \sigma_{-i})$. But if so, then there must be some pure strategy s_i' that is played with positive probability under σ_i' for which $u_i(s_i', \sigma_{-i}) > u_i(\sigma_i, \sigma_{-i})$. Since $u_i(\sigma_i, \sigma_{-i}) = u_i(s_i, \sigma_{-i})$ for all $s_i \in S_i^+$, this contradicts conditions (i) and (ii) being satisfied. ■

Hence, a necessary and sufficient condition for mixed strategy profile σ to be a Nash equilibrium of game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ is that each player, given the distribution of strategies played by his opponents, is indifferent among all the pure strategies that he plays with positive probability and that these pure strategies are at least as good as any pure strategy he plays with zero probability.

An implication of Proposition 8.D.1 is that to test whether a strategy profile σ is a Nash equilibrium it suffices to consider only pure strategy deviations (i.e., changes in a player's strategy σ_i to some pure strategy s_i'). As long as no player can improve his payoff by switching to any pure strategy, σ is a Nash equilibrium. We therefore get the comforting result given in Corollary 8.D.1.

Corollary 8.D.1: Pure strategy profile $s = (s_1, \ldots, s_I)$ is a Nash equilibrium of game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if and only if it is a (degenerate) mixed strategy Nash equilibrium of game $\Gamma_N' = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$.

Corollary 8.D.1 tells us that to identify the pure strategy equilibria of game $\Gamma_N' = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, it suffices to restrict attention to the game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ in which randomization is not permitted.

Proposition 8.D.1 can also be of great help in the computation of mixed strategy equilibria as Example 8.D.5 illustrates.

Example 8.D.5: *Mixed Strategy Equilibria in the Meeting in New York Game.* Let us try to find a mixed strategy equilibrium in the variation of the Meeting in New York game where the payoffs of meeting at Grand Central are (1000, 1000). By Proposition 8.D.1, if Mr. Thomas is going to randomize between Empire State and Grand Central, he must be indifferent between them. Suppose that Mr. Schelling plays Grand Central with probability σ_s. Then Mr. Thomas' expected payoff from playing Grand Central is $1000\sigma_s + 0(1 - \sigma_s)$, and his expected payoff from playing Empire State is $100(1 - \sigma_s) + 0\sigma_s$. These two expected payoffs are equal only when $\sigma_s = 1/11$. Now, for Mr. Schelling to set $\sigma_s = 1/11$, he must also be indifferent between his two pure strategies. By a similar argument, we find that Mr. Thomas' probability of playing Grand Central must also be 1/11. We conclude that each player going to Grand Central with a probability of 1/11 is a Nash equilibrium. ■

Note that in accordance with Proposition 8.D.1, the players in Example 8.D.5 have no real preference over the probabilities that they assign to the pure strategies they play with positive probability. What determines the probabilities that each player uses is an equilibrium consideration: the need to make the *other* player indifferent over *his* strategies.

This fact has led some economists and game theorists to question the usefulness of mixed strategy Nash equilibria as predictions of play. They raise two concerns: First, if players always have a pure strategy that gives them the same expected payoff as their equilibrium mixed strategy, it is not clear why they will bother to randomize.

One answer to this objection is that players may not actually randomize. Rather, they may make definite choices that are affected by seemingly inconsequential variables ("signals") that only they observe. For example, consider how a pitcher for a major league baseball team "mixes his pitches" to keep batters guessing. He may have a completely deterministic plan for what he will do, but it may depend on which side of the bed he woke up on that day or on the number of red traffic lights he came to on his drive to the stadium. As a result, batters view the behavior of the pitcher as random even though it is not. We touched briefly on this interpretation of mixed strategies as behavior contingent on realizations of a signal in Section 7.E, and we will examine it in more detail in Section 8.E.

The second concern is that the stability of mixed strategy equilibria seems tenuous. Players must randomize with exactly the correct probabilities, but they have no positive incentive to do so. One's reaction to this problem may depend on why one expects a Nash equilibrium to arise in the first place. For example, the use of the correct probabilities may be unlikely to arise as a stable social convention, but may seem more plausible when the equilibrium arises as a self-enforcing agreement.

Up to this point, we have assumed that players' randomizations are independent. In the Meeting in New York game in Example 8.D.5, for instance, we could describe a mixed strategy equilibrium as follows: Nature provides *private and independently distributed* signals $(\theta_1, \theta_2) \in [0, 1] \times [0, 1]$ to the two players, and each player i assigns decisions to the various possible realizations of his signal θ_i.

However, suppose that there are also *public* signals available that both players observe. Let $\theta \in [0, 1]$ be such a signal. Then many new possibilities arise. For example, the two players could both decide to go to Grand Central if $\theta < \frac{1}{2}$ and to Empire State if $\theta \geq \frac{1}{2}$. Each player's strategy choice is still random, but the coordination of their actions is now perfect and they always meet. More importantly, the decisions have an equilibrium character. If one player decides to follow this decision rule, then it is also optimal for the other player to do so. This is an example of a *correlated equilibrium* [due to Aumann (1974)]. More generally, we could allow for correlated equilibria in which nature's signals are partly private and partly public.

Allowing for such correlation may be important because economic agents observe many public signals. Formally, a correlated equilibrium is a special case of a Bayesian Nash equilibrium, a concept that we introduce in Section 8.E; hence, we defer further discussion to the end of that section.

Existence of Nash Equilibria

Does a Nash equilibrium necessarily exist in a game? Fortunately, the answer turns out to be "yes" under fairly broad circumstances. Here we describe two of the more important existence results; their proofs, based on mathematical fixed point theorems, are given in Appendix A of this chapter. (Proposition 9.B.1 of Section 9.B provides another existence result.)

Proposition 8.D.2: Every game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ in which the sets $S_1, \ldots, S_I$ have a finite number of elements has a mixed strategy Nash equilibrium.

Thus, for the class of games we have been considering, a Nash equilibrium always exists as long as we are willing to accept equilibria in which players randomize. (If you want to be convinced without going through the proof, try Exercise 8.D.6.) Allowing

maximization is essential for this result. We have already seen in (standard) Matching Pennies, for example, that a pure strategy equilibrium may not exist in a game with a finite number of pure strategies.

Up to this point, we have focused on games with finite strategy sets. However, in economic applications, we frequently encounter games in which players have strategies naturally modeled as continuous variables. This can be helpful for the existence of a pure strategy equilibrium. In particular, we have the result given in Proposition 8.D.3.

Proposition 8.D.3: A Nash equilibrium exists in game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if for all $i = 1, \ldots, I$,

(i) S_i is a nonempty, convex, and compact subset of some Euclidean space $\mathbb{R}^M$.

(ii) $u_i(s_1, \ldots, s_I)$ is continuous in $(s_1, \ldots, s_I)$ and quasiconcave in s_i.

Proposition 8.D.3 provides a significant result whose requirements are satisfied in a wide range of economic applications. The convexity of strategy sets and the nature of the payoff functions help to smooth out the structure of the model, allowing us to achieve a pure strategy equilibrium.[5]

Further existence results can also be established. In situations where quasiconcavity of the payoff functions $u_i(\cdot)$ fails but they are still continuous, existence of a mixed strategy equilibrium can still be demonstrated. In fact, even if continuity of the payoff functions fails to hold, a mixed strategy equilibrium can be shown to exist in a variety of cases [see Dasgupta and Maskin (1986)].

Of course, these results do not mean that we *cannot* have an equilibrium if the conditions of these existence results do not hold. Rather, we just cannot be *assured* that there is one.

8.E Games of Incomplete Information: Bayesian Nash Equilibrium

Up to this point, we have assumed that players know all relevant information about each other, including the payoffs that each receives from the various outcomes of the game. Such games are known as games of *complete information.* A moment of thought, however, should convince you that this is a very strong assumption. Do two firms in an industry necessarily know each other's costs? Does a firm bargaining with a union necessarily know the disutility that union members will feel if they go out on strike for a month? Clearly, the answer is "no." Rather, in many circumstances, players have what is known as *incomplete information.*

The presence of incomplete information raises the possibility that we may need to consider a player's beliefs about other players' preferences, his beliefs about their beliefs about his preferences, and so on, much in the spirit of rationalizability.[6]

---

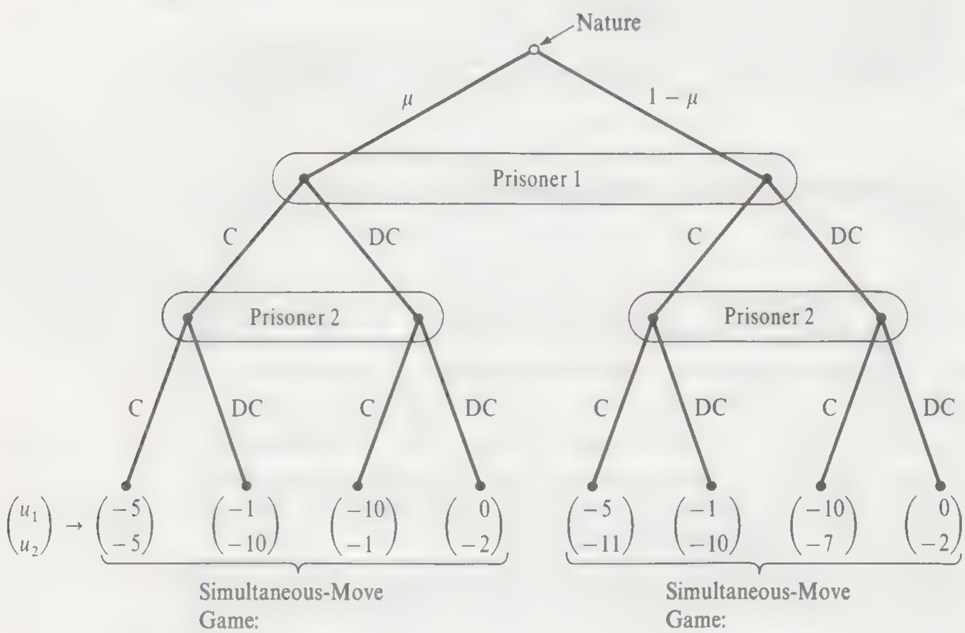
5. Note that a finite strategy set S_i cannot be convex. In fact, the use of mixed strategies in Proposition 8.D.2 helps us to obtain existence of equilibrium in much the same way that Proposition 8.D.3's assumptions assure existence of a pure strategy Nash equilibrium: It convexifies players' strategy sets and yields well-behaved payoff functions. (See Appendix A for details.)

6. For more on this problem, see Mertens and Zamir (1985).

Fortunately, there is a widely used approach to this problem, originated by Harsanyi (1967-68), that makes this unnecessary. In this approach, one imagines that each player's preferences are determined by the realization of a random variable. Although the random variable's actual realization is observed only by the player, its ex ante probability distribution is assumed to be common knowledge among all the players. Through this formulation, the situation of incomplete information is reinterpreted as a game of imperfect information: Nature makes the first move, choosing realizations of the random variables that determine each player's preference *type*, and each player observes the realization of only his own random variable. A game of this sort is known as a *Bayesian game.*

Example 8.E.1: Consider a modification of the DA's Brother game discussed in Example 8.B.3. With probability μ, prisoner 2 has the preferences in Figure 8.B.4 (we call these *type I preferences*), while with probability $(1 - \mu)$, prisoner 2 hates to rat on his accomplice (this is *type II*). In this case, he pays a psychic penalty equal to 6 years in prison for confessing. Prisoner 1, on the other hand, always has the preferences depicted in Figure 8.B.4. The extensive form of this Bayesian game is represented in Figure 8.E.1 (in the figure, "C" and "DC" stand for "confess" and "don't confess" respectively).

In this game, a pure strategy (a complete contingent plan) for player 2 can be viewed as a function that for each possible realization of his preference type



Prisoner 1 \ Prisoner 2	DC	C
DC	0, −2	−10, −1
C	−1, −10	−5, −5

Prisoner 1 \ Prisoner 2	DC	C
DC	0, −2	−10, −7
C	−1, −10	−5, −11

Figure 8.E.1 The DA's B[illegible] game with [illegible] information.

ates what action he will take. Hence, prisoner 2 now has four possible strategies:

confess if type I, confess if type II);
confess if type I, don't confess if type II);
don't confess if type I, confess if type II);
don't confess if type I, don't confess if type II).

e, however, that player 1 does not observe player 2's type, and so a pure egy for player 1 in this game is simply a (noncontingent) choice of either ess" or "don't confess." ∎

mally, in a Bayesian game, each player i has a payoff function $u_i(s_i, s_{-i}, \theta_i)$, $\in \Theta_i$ is a random variable chosen by nature that is observed only by player joint probability distribution of the θ_i's is given by $F(\theta_1, \dots, \theta_I)$, which is ed to be common knowledge among the players. Letting $\Theta = \Theta_1 \times \cdots \times \Theta_I$, yesian game is summarized by the data $[I, \{S_i\}, \{u_i(\cdot)\}, \Theta, F(\cdot)]$.

pure strategy for player i in a Bayesian game is a function $s_i(\theta_i)$, or *decision* hat gives the player's strategy choice for each realization of his type θ_i. Player re strategy set $\mathcal{S}_i$ is therefore the set of all such functions. Player i's expected off given a profile of pure strategies for the I players $(s_1(\cdot), \dots, s_I(\cdot))$ is then by

$$\tilde{u}_i(s_1(\cdot), \dots, s_I(\cdot)) = E_\theta[u_i(s_1(\theta_1), \dots, s_I(\theta_I), \theta_i)]. \tag{8.E.1}$$

e can now look for an ordinary (pure strategy) Nash equilibrium of this game mperfect information, which is known in this context as a *Bayesian Nash* rium.[7]

n 8.E.1: A (pure strategy) *Bayesian Nash equilibrium* for the Bayesian game $S_i\}, \{u_i(\cdot)\}, \Theta, F(\cdot)]$ is a profile of decision rules $(s_1(\cdot), \dots, s_I(\cdot))$ that con-tutes a Nash equilibrium of game $\Gamma_N = [I, \{\mathcal{S}_i\}, \{\tilde{u}_i(\cdot)\}]$. That is, for every $\dots, I$,

$$\tilde{u}_i(s_i(\cdot), s_{-i}(\cdot)) \geq \tilde{u}_i(s_i'(\cdot), s_{-i}(\cdot))$$

all $s_i'(\cdot) \in \mathcal{S}_i$, where $\tilde{u}_i(s_i(\cdot), s_{-i}(\cdot))$ is defined as in (8.E.1).

very useful point to note is that in a (pure strategy) Bayesian Nash equilibrium player must be playing a best response to the conditional distribution of his nents' strategies *for each type that he might end up having.* Proposition 8.E.1 vides a more formal statement of this point.

on 8.E.1: A profile of decision rules $(s_1(\cdot), \dots, s_I(\cdot))$ is a Bayesian Nash orium in Bayesian game $[I, \{S_i\}, \{u_i(\cdot)\}, \Theta, F(\cdot)]$ if and only if, for all i and

7. We shall restrict our attention to pure strategies here; mixed strategies involve randomization the strategies in $\mathcal{S}_i$. Note also that we have not been very explicit about whether the Θ_i's are sets. If they are, then the strategy sets $\mathcal{S}_i$ are finite; if they are not, then the sets $\mathcal{S}_i$ include an te number of possible functions $s_i(\cdot)$. Either way, however, the basic definition of a Bayesian equilibrium is the same.

all $\bar{\theta}_i \in \Theta_i$ occurring with positive probability[8]

$$E_{\theta_{-i}}[u_i(s_i(\bar{\theta}_i), s_{-i}(\theta_{-i}), \bar{\theta}_i) | \bar{\theta}_i] \geq E_{\theta_{-i}}[u_i(s_i', s_{-i}(\theta_{-i}), \bar{\theta}_i) | \bar{\theta}_i] \tag{8.E.2}$$

for all $s_i' \in S_i$, where the expectation is taken over realizations of the other players' random variables conditional on player i's realization of his signal $\bar{\theta}_i$.

Proof: For necessity, note that if (8.E.2) did not hold for some player i for some $\bar{\theta}_i \in \Theta_i$ that occurs with positive probability, then player i could do better by changing his strategy choice in the event he gets realization $\bar{\theta}_i$, contradicting $(s_1(\cdot), \ldots, s_I(\cdot))$ being a Bayesian Nash equilibrium. In the other direction, if condition (8.E.2) holds for all $\bar{\theta}_i \in \Theta_i$ occurring with positive probability, then player i cannot improve on the payoff he receives by playing strategy $s_i(\cdot)$. ■

Proposition 8.E.1 tells us that, in essence, we can think of each type of player i as being a separate player who maximizes his payoff given his conditional probability distribution over the strategy choices of his rivals.

Example 8.E.1 Continued: To solve for the (pure strategy) Bayesian Nash equilibrium of this game, note first that type I of prisoner 2 must play "confess" with probability 1 because this is that type's dominant strategy. Likewise, type II of prisoner 2 also has a dominant strategy: "don't confess." Given this behavior by prisoner 2, prisoner 1's best response is to play "don't confess" if $[-10\mu + 0(1-\mu)] > [-5\mu - 1(1-\mu)]$, or equivalently, if $\mu < \frac{1}{6}$, and is to play "confess" if $\mu > \frac{1}{6}$. (He is indifferent if $\mu = \frac{1}{6}$.) ■

Example 8.E.2: The Alphabeta research and development consortium has two (noncompeting) members, firms 1 and 2. The rules of the consortium are that any independent invention by one of the firms is shared fully with the other. Suppose that there is a new invention, the "Zigger," that either of the two firms could potentially develop. To develop this new product costs a firm $c \in (0, 1)$. The benefit of the Zigger to each firm i is known only by that firm. Formally, each firm i has a type θ_i that is independently drawn from a uniform distribution on $[0, 1]$, and its benefit from the Zigger when its type is θ_i is $(\theta_i)^2$. The timing is as follows: The two firms each privately observe their own type. Then they each simultaneously choose either to develop the Zigger or not.

Let us now solve for the Bayesian Nash equilibrium of this game. We shall write $s_i(\theta_i) = 1$ if type θ_i of firm i develops the Zigger and $s_i(\theta_i) = 0$ if it does not. If firm i develops the Zigger when its type is θ_i, its payoff is $(\theta_i)^2 - c$ regardless of whether firm j does so. If firm i decides not to develop the Zigger when its type is θ_i, it will have an expected payoff equal to $(\theta_i)^2 \text{ Prob } (s_j(\theta_j) = 1)$. Hence, firm i's best response is to develop the Zigger if and only if its type θ_i is such that (we assume firm i develops the Zigger if it is indifferent):

$$\theta_i \geq \left[\frac{c}{1 - \text{Prob }(s_j(\theta_j) = 1)}\right]^{1/2}. \tag{8.E.3}$$

---

8. The formulation given here (and the proof) is for the case in which the sets Θ_i are finite. When a player i has an infinite number of possible types, condition (8.E.2) must hold on a subset of Θ_i that is of full measure (i.e., that occurs with probability equal to one). It is then said that (8.E.2) holds for *almost every* $\bar{\theta}_i \in \Theta_i$.

Note that for any given strategy of firm j, firm i's best response takes the form of a *cutoff rule*: It optimally develops the Zigger for all θ_i above the value on the right-hand side of (8.E.3) and does not for all θ_i below it. [Note that if firm i existed in isolation, it would be indifferent about developing the Zigger when $\theta_i = \sqrt{c}$. But (8.E.3) tells us that when firm i is part of the consortium, its cutoff is always (weakly) above this. This is true because each firm hopes to *free-ride* on the other firm's development effort; see Chapter 11 for more on this.]

Suppose then that $\hat{\theta}_1, \hat{\theta}_2 \in (0, 1)$ are the cutoff values for firms 1 and 2 respectively in a Bayesian Nash equilibrium (it can be shown that $0 < \hat{\theta}_i < 1$ for $i = 1, 2$ in any Bayesian Nash equilibrium of this game). If so, then using the fact that $\text{Prob}(s_j(\theta_j) = 1) = 1 - \hat{\theta}_j$, condition (8.E.3) applied first for $i = 1$ and then for $i = 2$ tells us that we must have

$$(\hat{\theta}_1)^2\hat{\theta}_2 = c$$

and

$$(\hat{\theta}_2)^2\hat{\theta}_1 = c.$$

Because $(\hat{\theta}_1)^2\hat{\theta}_2 = (\hat{\theta}_2)^2\hat{\theta}_1$ implies that $\hat{\theta}_1 = \hat{\theta}_2$, we see that any Bayesian Nash equilibrium of this game involves an identical cutoff value for the two firms, $\theta^* = (c)^{1/3}$. In this equilibrium, the probability that neither firm develops the Zigger is $(\theta^*)^2$, the probability that exactly one firm develops it is $2\theta^*(1 - \theta^*)$, and the probability that both do is $(1 - \theta^*)^2$. ■

The exercises at the end of this chapter consider several other examples of Bayesian Nash equilibria. Another important application arises in the theory of implementation with incomplete information, studied in Chapter 23.

In Section 8.D, we argued that mixed strategies could be interpreted as situations where players play deterministic strategies conditional on seemingly irrelevant signals (recall the baseball pitcher). We can now say a bit more about this. Suppose we start with a game of complete information that has a unique mixed strategy equilibrium in which players actually randomize. Now consider changing the game by introducing many different types (formally, a continuum) of each player, with the realizations of the various players' types being statistically independent of one another. Suppose, in addition, that all types of a player have *identical* preferences. A (pure strategy) Bayesian Nash equilibrium of this Bayesian game is then precisely equivalent to a mixed strategy Nash equilibrium of the original complete information game. Moreover, in many circumstances, one can show that there are also "nearby" Bayesian games in which preferences of the different types of a player differ only slightly from one another, the Bayesian Nash equilibria are close to the mixed strategy combination, and each type has a strict preference for his strategy choice. Such results are known as *purification theorems* [see Harsanyi (1973)].

We can also return to the issue of *correlated equilibria* raised in Section 8.D. In particular, if we allow the realizations of the various players' types in the previous paragraph to be arbitrarily correlated, then a (pure strategy) Bayesian Nash equilibrium of this Bayesian game is a correlated equilibrium of the original complete information game. The set of all correlated equilibria in game $[I, \{S_i\}, \{u_i(\cdot)\}]$ is identified by considering all possible Bayesian games of this sort (i.e., we allow for all possible signals that the players might observe).

8.F The Possibility of Mistakes: Trembling-Hand Perfection

In Section 8.B, we noted that although rationality per se does not rule out the choice of a weakly dominated strategy, such strategies are unappealing because they are dominated unless a player is absolutely sure of what his rivals will play. In fact, as the game depicted in Figure 8.F.1 illustrates, the Nash equilibrium concept also does not preclude the use of such strategies. In this game, (D, R) is a Nash equilibrium in which both players play a weakly dominated strategy with certainty.

Here, we elaborate on the idea, raised in Section 8.B, that *caution* might preclude the use of such strategies. The discussion leads us to define a refinement of the concept of Nash equilibrium, known as a (*normal form*) *trembling-hand perfect Nash equilibrium*, which identifies Nash equilibria that are robust to the possibility that, with some very small probability, players make mistakes.

Following Selten (1975), for any normal form game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, we can define a *perturbed* game $\Gamma_\varepsilon = [I, \{\Delta_\varepsilon(S_i)\}, \{u_i(\cdot)\}]$ by choosing for each player i and strategy $s_i \in S_i$ a number $\varepsilon_i(s_i) \in (0, 1)$, with $\sum_{s_i \in S_i} \varepsilon_i(s_i) < 1$, and then defining player i's perturbed strategy set to be

$$\Delta_\varepsilon(S_i) = \{\sigma_i: \sigma_i(s_i) \geq \varepsilon_i(s_i) \text{ for all } s_i \in S_i \text{ and } \sum_{s_i \in S_i} \sigma_i(s_i) = 1\}.$$

That is, perturbed game Γ_ε is derived from the original game Γ_N by requiring that each player i play every one of his strategies, say s_i, with at least some minimal positive probability $\varepsilon_i(s_i)$; $\varepsilon_i(s_i)$ is interpreted as the unavoidable probability that strategy s_i gets played by mistake.

Having defined this perturbed game, we want to consider as predictions in game Γ_N only those Nash equilibria σ that are robust to the possibility that players make mistakes. The robustness test we employ can be stated roughly as: To consider σ as a robust equilibrium, we want there to be at least some slight perturbations of Γ_N whose equilibria are close to σ. The formal definition of a (*normal form*) *trembling-hand perfect Nash equilibrium* (the name comes from the idea of players making mistakes because of their trembling hands) is presented in Definition 8.F.1.

Definition 8.F.1: A Nash equilibrium σ of game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ is (*normal form*) *trembling-hand perfect* if there is *some* sequence of perturbed games $\{\Gamma_{\varepsilon^k}\}_{k=1}^{\infty}$ that converges to Γ_N [in the sense that $\lim_{k\to\infty} \varepsilon_i^k(s_i) = 0$ for all i and $s_i \in S_i$], for which there is *some* associated sequence of Nash equilibria $\{\sigma^k\}_{k=1}^{\infty}$ that converges to σ (i.e., such that $\lim_{k\to\infty} \sigma^k = \sigma$).

We use the modifier *normal form* because Selten (1975) also proposes a slightly different form of trembling-hand perfection for dynamic games; we discuss this version of the concept in Chapter 9.[9]

Note that the concept of a (normal form) trembling-hand perfect Nash equilibrium provides a relatively mild test of robustness: We require only that *some* perturbed games exist that have equilibria arbitrarily close to σ. A stronger test would

9. In fact, Selten (1975) is primarily concerned with the problem of identifying desirable equilibria in dynamic games. See Chapter 9, Appendix B for more on this.

	L	R
U	1, 1	0, −3
D	−3, 0	0, 0

Figure 8.F.1
(D, R) is a Nash equilibrium involving play of weakly dominated strategies.

ure that the equilibrium σ be robust to *all* perturbations close to the original game.

general, the criterion in Definition 8.F.1 can be difficult to work with because it s that we compute the equilibria of many possible perturbed games. The result ed in Proposition 8.F.1 provides a formulation that makes checking whether a equilibrium is trembling-hand perfect much easier (in its statement, a *totally* strategy is a mixed strategy in which every pure strategy receives positive ability).

on 8.F.1: A Nash equilibrium σ of game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ is (normal embling-hand perfect if and only if there is some sequence of totally mixed gies $\{\sigma^k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} \sigma^k = \sigma$ and σ_i is a best response to every ent of sequence $\{\sigma_{-i}^k\}_{k=1}^{\infty}$ for all $i = 1, \ldots, I$.

are asked to prove this result in Exercise 8.F.1 [or consult Selten (1975)]. The esented in Proposition 8.F.2 is an immediate consequence of Definition 8.F.1 Proposition 8.F.1.

on 8.F.2: If $\sigma = (\sigma_1, \ldots, \sigma_I)$ is a (normal form) trembling-hand perfect Nash brium, then σ_i is not a weakly dominated strategy for any $i = 1, \ldots, I$. any (normal form) trembling-hand perfect Nash equilibrium, no weakly ated pure strategy can be played with positive probability.

---

e converse, that any Nash equilibrium not involving play of a weakly dominated strategy ssarily trembling-hand perfect, turns out to be true for two-player games but not for with more than two players. Thus, trembling-hand perfection can rule out more than sh equilibria involving weakly dominated strategies. The reason is tied to the fact that a player's rivals make mistakes with small probability, this can give rise to only a limited probability distributions over their nonequilibrium strategies. For example, if a player's ls each have a small probability of making a mistake, there is a much greater ty that one will make a mistake than that both will. If the player's equilibrium strategy que best response only when both of his rivals make a mistake, his strategy may not est response to any local perturbation of his rivals' strategies even though his strategy eakly dominated. (Exercise 8.F.2 provides an example.) However, if players' trembles ed to be correlated (e.g., as in the correlated equilibrium concept), then the converse sition 8.F.2 would hold regardless of the number of players.

---

en (1975) also proves an existence result that parallels Proposition 8.D.2: game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ with finite strategy sets $S_1, \ldots, S_I$ has a g-hand perfect Nash equilibrium. An implication of this result is that every me has at least one Nash equilibrium in which no player plays any weakly strategy with positive probability. Hence, if we decide to accept only Nash

equilibria that do not involve the play of weakly dominated strategies, with great generality there is at least one such equilibrium.[10]

Myerson (1978) proposes a refinement of Selten's idea in which players are less likely to make more costly mistakes (the idea is that they will try harder to avoid these mistakes). He establishes that the resulting solution concept, called a *proper Nash equilibrium*, exists under the conditions described in the previous paragraph for trembling-hand perfect Nash equilibria. van Damme (1983) presents a good discussion of this and other refinements of trembling-hand perfection.

APPENDIX A: EXISTENCE OF NASH EQUILIBRIUM

In this appendix, we prove Propositions 8.D.2 and 8.D.3. We begin with Lemma 8.AA.1, which provides a key technical result.

Lemma 8.AA.1: If the sets $S_1, \ldots, S_I$ are nonempty, S_i is compact and convex, and $u_i(\cdot)$ is continuous in $(s_1, \ldots, s_I)$ and quasiconcave in s_i, then player i's best-response correspondence $b_i(\cdot)$ is nonempty, convex-valued, and upper hemicontinuous.[11]

Proof: Note first that $b_i(s_{-i})$ is the set of maximizers of the continuous function $u_i(\cdot, s_{-i})$ on the compact set S_i. Hence, it is nonempty (see Theorem M.F.2 of the Mathematical Appendix). The convexity of $b_i(s_{-i})$ follows because the set of maximizers of a quasiconcave function [here, the function $u_i(\cdot, s_{-i})$] on a convex set (here, S_i) is convex. Finally, for upper hemicontinuity, we need to show that for any sequence $(s_i^n, s_{-i}^n) \to (s_i, s_{-i})$ such that $s_i^n \in b_i(s_{-i}^n)$ for all n, we have $s_i \in b_i(s_{-i})$. To see this, note that for all n, $u_i(s_i^n, s_{-i}^n) \geq u_i(s_i', s_{-i}^n)$ for all $s_i' \in S_i$. Therefore, by the continuity of $u_i(\cdot)$, we have $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$. ■

It is convenient to prove Proposition 8.D.3 first.

Proposition 8.D.3: A Nash equilibrium exists in game $\Gamma_N = [I, \{S_i\}, \{u_i(\cdot)\}]$ if for all $i = 1, \ldots, I$,

(i) S_i is a nonempty, convex, and compact subset of some Euclidean space $\mathbb{R}^M$.

(ii) $u_i(s_1, \ldots, s_I)$ *is continuous in* $(s_1, \ldots, s_I)$ and quasiconcave in s_i.

---

10. The Bertrand duopoly game discussed in Chapter 12 provides one example of a game in which this is not the case; its unique Nash equilibrium involves the play of weakly dominated strategies. The problem arises because the strategies in that game are continuous variables (and so the sets S_i are not finite). Fortunately, this equilibrium can be viewed as the limit of undominated equilibria in "nearby" discrete versions of the game. (See Exercise 12.C.3 for more on this point.)

11. See Section M.H of the Mathematical Appendix for a discussion of upper hemicontinuous correspondences.

Define the correspondence $b: S \to S$ by

$$b(s_1, \ldots, s_I) = b_1(s_{-1}) \times \cdots \times b_I(s_{-I}).$$

that $b(\cdot)$ is a correspondence from the nonempty, convex, and compact set $S_1 \times \cdots \times S_I$ to itself. In addition, by Lemma 8.AA.1, $b(\cdot)$ is a nonempty, -valued, and upper hemicontinuous correspondence. Thus, all the conditions akutani fixed point theorem are satisfied (see Section M.I of the Mathematical). Hence, there exists a fixed point for this correspondence, a strategy profile ch that $s \in b(s)$. The strategies at this fixed point constitute a Nash equilibrium se by construction $s_i \in b_i(s_{-i})$ for all $i = 1, \ldots, I$. ■

ow we move to the proof of Proposition 8.D.2.

on 8.D.2: Every game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ in which the sets $S_1, \ldots, S_I$ a finite number of elements has a mixed strategy Nash equilibrium.

The game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$, viewed as a game with strategy sets and payoff functions $u_i(\sigma_1, \ldots, \sigma_I) = \sum_{s \in S}[\prod_{k=1}^{I} \sigma_k(s_k)]u_i(s)$ for all $i = 1, \ldots, I$, all the assumptions of Proposition 8.D.3. Hence, Proposition 8.D.2 is a direct ary of that result. ■

ERENCES

R. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics* **1**: 67–96.

B. D. (1984). Rationalizable strategic behavior. *Econometrica* **52**: 1007–28.

B. D. (1986). Axiomatic characterizations of rational choice in strategic environments. *Scandinavian Journal of Economics* **88**: 473–88.

berger, A., and E. Dekel. (1987). Rationalizability and correlated equilibria. *Econometrica* **55**: 1391–1402.

A. (1838). *Recherches sur les Principes Mathématiques de la Théorie des Richesses.* [English edition: *Researches into the Mathematical Principles of the Theory of Wealth.* New York: Macmillan, 1897.]

P., and E. Maskin. (1986). The existence of equilibrium in discontinuous economic games. *Review of Economic Studies* **53**: 1–41.

yi, J. (1967–68). Games with incomplete information played by Bayesian players. *Management Science* **14**: 159–82, 320–34, 486–502.

yi, J. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory* **2**: 1–23.

D. M. (1990). *Game Theory and Economic Modelling.* Oxford: Oxford University Press.

s, J. F., and S. Zamir. (1985). Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* **10**: 619–32.

m, P., and J. Roberts. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica* **58**: 1255–78.

on, R. B. (1978). Refinements of the Nash equilibrium concept. *International Journal of Game Theory* **7**: 73–80.

J. F. (1951). Non-cooperative games. *Annals of Mathematics* **54**: 289–95.

D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica* **52**: 1029–50.

T. (1960). *The Strategy of Conflict.* Cambridge, Mass.: Harvard University Press.

R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**: 25–55.

Damme, E. (1983). *Refinements of the Nash Equilibrium Concept.* Berlin: Springer-Verlag.

EXERCISES

8.B.1[A] There are I firms in an industry. Each can try to convince Congress to give the industry a subsidy. Let h_i denote the number of hours of effort put in by firm i, and let $c_i(h_i) = w_i(h_i)^2$, where w_i is a positive constant, be the cost of this effort to firm i. When the effort levels of the firms are $(h_1, \ldots, h_I)$, the value of the subsidy that gets approved is $\alpha\sum_i h_i + \beta(\prod_i h_i)$, where α and β are constants.

Consider a game in which the firms decide simultaneously and independently how many hours they will each devote to this effort. Show that each firm has a strictly dominant strategy if and only if $\beta = 0$. What is firm i's strictly dominant strategy when this is so?

8.B.2[B] **(a)** Argue that if a player has two weakly dominant strategies, then for every strategy choice by his opponents, the two strategies yield him equal payoffs.

(b) Provide an example of a two-player game in which a player has two weakly dominant pure strategies but his opponent prefers that he play one of them rather than the other.

8.B.3[B] Consider the following auction (known as a *second-price*, or *Vickrey*, auction). An object is auctioned off to I bidders. Bidder i's valuation of the object (in monetary terms) is v_i. The auction rules are that each player submit a bid (a nonnegative number) in a sealed envelope. The envelopes are then opened, and the bidder who has submitted the highest bid gets the object but pays the auctioneer the amount of the *second-highest* bid. If more than one bidder submits the highest bid, each gets the object with equal probability. Show that submitting a bid of v_i with certainty is a weakly dominant strategy for bidder i. Also argue that this is bidder i's unique weakly dominant strategy.

8.B.4[C] Show that the order of deletion does not matter for the set of strategies surviving a process of iterated deletion of strictly dominated strategies.

8.B.5[C] Consider the Cournot duopoly model (discussed extensively in Chapter 12) in which two firms, 1 and 2, simultaneously choose the quantities they will sell on the market, q_1 and q_2. The price each receives for each unit given these quantities is $P(q_1, q_2) = a - b(q_1 + q_2)$. Their costs are c per unit sold.

(a) Argue that successive elimination of strictly dominated strategies yields a unique prediction in this game.

(b) Would this be true if there were three firms instead of two?

8.B.6[B] In text.

8.B.7[B] Show that any strictly dominant strategy in game $[I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ must be a pure strategy.

8.C.1[A] Argue that if elimination of strictly dominated strategies yields a unique prediction in a game, this prediction also results from eliminating strategies that are never a best response.

8.C.2[C] Prove that the order of removal does not matter for the set of strategies that survives a process of iterated deletion of strategies that are never a best response.

8.C.3[C] Prove that in a two-player game (with finite strategy sets), if a pure strategy s_i for player i is never a best response for any mixed strategy by i's opponent, then s_i is strictly dominated by some mixed strategy $\sigma_i \in \Delta(S_i)$. [*Hint:* Try using the supporting hyperplane theorem presented in Section M.G of the Mathematical Appendix.]

Consider a game Γ_N with players 1, 2, and 3 in which $S_1 = \{L, M, R\}$, $S_2 = \{U, D\}$, $S_3 = \{\ell, r\}$. Player 1's payoffs from each of his three strategies conditional on the strategy ...es of players 2 and 3 are depicted as (u_L, u_M, u_R) in each of the four boxes shown below, ...re $(\pi, \varepsilon, \eta) \gg 0$. Assume that $\eta < 4\varepsilon$.

Player 3's Strategy

Player 2's Strategy		ℓ	r
	U	$\pi + 4\varepsilon, \pi - \eta, \pi - 4\varepsilon$	$\pi - 4\varepsilon, \pi + \dfrac{\eta}{2}, \pi + 4\varepsilon$
	D	$\eta + 4\varepsilon, \pi + \dfrac{\eta}{2}, \pi - 4\varepsilon$	$\pi - 4\varepsilon, \pi - \eta, \pi + 4\varepsilon$

(a) Argue that (pure) strategy M is never a best response for player 1 to any independent ...omizations by players 2 and 3.

(b) Show that (pure) strategy M is not strictly dominated.

(c) Show that (pure) strategy M can be a best response if player 2's and player 3's ...omizations are allowed to be correlated.

Show that (a_2, b_2) being played with certainty is the unique mixed strategy Nash ...rium in the game depicted in Figure 8.C.1.

Show that if there is a unique profile of strategies that survives iterated removal of ... dominated strategies, this profile is a Nash equilibrium.

Consider a first-price sealed-bid auction of an object with two bidders. Each bidder ...aluation of the object is v_i, which is known to both bidders. The auction rules are that ...ayer submits a bid in a sealed envelope. The envelopes are then opened, and the bidder ... submitted the highest bid gets the object and pays the auctioneer the amount of his ... the bidders submit the same bid, each gets the object with probability $\frac{1}{2}$. Bids must be ...llar multiples (assume that valuations are also).

(a) Are any strategies strictly dominated?

(b) Are any strategies weakly dominated?

(c) Is there a Nash equilibrium? What is it? Is it unique?

Consider a bargaining situation in which two individuals are considering undertaking ...siness venture that will earn them 100 dollars in profit, but they must agree on how to ... the 100 dollars. Bargaining works as follows: The two individuals each make a demand ...ltaneously. If their demands sum to more than 100 dollars, then they fail to agree, and ... gets nothing. If their demands sum to less than 100 dollars, they do the project, each gets his demand, and the rest goes to charity.

(a) What are each player's strictly dominated strategies?

(b) What are each player's weakly dominated strategies?

(c) What are the pure strategy Nash equilibria of this game?

Consumers are uniformly distributed along a boardwalk that is 1 mile long. Ice-cream ...ices are regulated, so consumers go to the nearest vendor because they dislike walking ...ssume that at the regulated prices all consumers will purchase an ice cream even if they

have to walk a full mile). If more than one vendor is at the same location, they split the business evenly.

(a) Consider a game in which two ice-cream vendors pick their locations simultaneously. Show that there exists a unique pure strategy Nash equilibrium and that it involves both vendors locating at the midpoint of the boardwalk.

(b) Show that with three vendors, no pure strategy Nash equilibrium exists.

8.D.6^B Consider any two-player game of the following form (where letters indicate arbitrary payoffs):

		Player 2	
		b_1	b_2
Player 1	a_1	u, v	ℓ, m
	a_2	w, x	y, z

Show that a mixed strategy Nash equilibrium always exists in this game. [*Hint:* Define player 1's strategy to be his probability of choosing action a_1 and player 2's to be his probability of choosing b_1; then examine the best-response correspondences of the two players.]

8.D.7^C (*The Minimax Theorem*) A two-player game with finite strategy sets $\Gamma_N = [I, \{S_1, S_2\}, \{u_1(\cdot), u_2(\cdot)\}]$ is a *zero-sum* game if $u_2(s_1, s_2) = -u_1(s_1, s_2)$ for all $(s_1, s_2) \in S_1 \times S_2$.

Define i's *maximin* expected utility level $\underline{w}_i$ to be the level he can guarantee himself in game $[I, \{\Delta(S_1), \Delta(S_2)\}, \{u_1(\cdot), u_2(\cdot)\}]$:

$$\underline{w}_i = \operatorname*{Max}_{\sigma_i}\left[\operatorname*{Min}_{\sigma_{-i}} u_i(\sigma_i, \sigma_{-i})\right].$$

Define player i's *minimax* utility level $\underline{v}_i$ to be the worst expected utility level he can be forced to receive if he gets to respond to his rival's actions:

$$\underline{v}_i = \operatorname*{Min}_{\sigma_{-i}}\left[\operatorname*{Max}_{\sigma_i} u_i(\sigma_i, \sigma_{-i})\right].$$

(a) Show that $\underline{v}_i \geq \underline{w}_i$ in any game.

(b) Prove that in any mixed strategy Nash equilibrium of the zero-sum game $\Gamma_N = [I, \{\Delta(S_1), \Delta(S_2)\}, \{u_1(\cdot), u_2(\cdot)\}]$, player i's expected utility u_i° satisfies $u_i^\circ = \underline{v}_i = \underline{w}_i$. [*Hint:* Such an equilibrium must exist by Proposition 8.D.2.]

(c) Show that if (σ_1', σ_2') and (σ_1'', σ_2'') are both Nash equilibria of the zero-sum game $\Gamma_N = [I, \{\Delta(S_1), \Delta(S_2)\}, \{u_1(\cdot), u_2(\cdot)\}]$, then so are (σ_1', σ_2'') and (σ_1'', σ_2').

8.D.8^C Consider a simultaneous-move game with normal form $[I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$. Suppose that, for all i, S_i is a convex set and $u_i(\cdot)$ is strictly quasiconvex. Argue that any mixed strategy Nash equilibrium of this game must be degenerate, with each player playing a single pure strategy with probability 1.

8.D.9^B Consider the following game [based on an example from Kreps (1990)]:

		Player 2			
		LL	L	M	R
Player 1	U	100, 2	-100, 1	0, 0	-100, -100
	D	-100, -100	100, -49	1, 0	100, 2

(a) If you were player 2 in this game and you were playing it once without the ability to engage in preplay communication with player 1, what strategy would you choose?

(b) What are all the Nash equilibria (pure and mixed) of this game?

(c) Is your strategy choice in **(a)** a component of any Nash equilibrium strategy profile? Is it a rationalizable strategy?

(d) Suppose now that preplay communication were possible. Would you expect to play anything different from your choice in **(a)**?

8.E.1[B] Consider the following strategic situation. Two opposed armies are poised to seize an island. Each army's general can choose either "attack" or "not attack." In addition, each army is either "strong" or "weak" with equal probability (the draws for each army are independent), and an army's type is known only to its general. Payoffs are as follows: The island is worth M if captured. An army can capture the island either by attacking when its opponent does not or by attacking when its rival does if it is strong and its rival is weak. If two armies of equal strength both attack, neither captures the island. An army also has a "cost" of fighting, which is s if it is strong and w if it is weak, where $s < w$. There is no cost of attacking if its rival does not.

Identify all pure strategy Bayesian Nash equilibria of this game.

8.E.2[C] Consider the first-price sealed-bid auction of Exercise 8.D.3, but now suppose that each bidder i observes only his own valuation v_i. This valuation is distributed uniformly and independently on $[0, \bar{v}]$ for each bidder.

(a) Derive a symmetric (pure strategy) Bayesian Nash equilibrium of this auction. (You should now suppose that bids can be any real number.) [*Hint:* Look for an equilibrium in which bidder i's bid is a linear function of his valuation.]

(b) What if there are I bidders? What happens to each bidder's equilibrium bid function $b(v_i)$ as I increases?

8.E.3[B] Consider the linear Cournot model described in Exercise 8.B.5. Now, however, suppose that each firm has probability μ of having unit costs of c_L and $(1-\mu)$ of having unit costs of c_H, where $c_H > c_L$. Solve for the Bayesian Nash equilibrium.

8.F.1[C] Prove Proposition 8.F.1.

8.F.2[B] Consider the following three-player game [taken from van Damme (1983)], in which player 1 chooses rows ($S_1 = \{U, D\}$), player 2 chooses columns ($S_2 = \{L, R\}$), and player 3 chooses boxes ($S_3 = \{B_1, B_2\}$):

B_1

	L	R
U	(1, 1, 1)	(1, 0, 1)
D	(1, 1, 1)	(0, 0, 1)

B_2

	L	R
U	(1, 1, 0)	(0, 0, 0)
D	(0, 1, 0)	(1, 0, 0)

Each cell describes the payoffs to the three players (u_1, u_2, u_3) from that strategy combination. Both (D, L, B_1) and (U, L, B_1) are pure strategy Nash equilibria. Show that (D, L, B_1) is not (normal form) trembling-hand perfect even though none of these three strategies is weakly dominated.

8.F.3^{C} Prove that every game $\Gamma_N = [I, \{\Delta(S_i)\}, \{u_i(\cdot)\}]$ in which the S_i are finite sets has a (normal form) trembling-hand perfect Nash equilibrium. [*Hint:* Show that every perturbed game has an equilibrium and that for any sequence of perturbed games converging to the original game Γ_N and corresponding sequence of equilibria, there is a subsequence that converges to an equilibrium of Γ_N.]

CHAPTER

9 Dynamic Games

9.A Introduction

In Chapter 8, we studied simultaneous-move games. Most economic situations, however, involve players choosing actions over time.[1] For example, a labor union and a firm might make repeated offers and counteroffers to each other in the course of negotiations over a new contract. Likewise, firms in a market may invest today in anticipation of the effects of these investments on their competitive interactions in the future. In this chapter, we therefore shift our focus to the study of *dynamic games*.

One way to approach the problem of prediction in dynamic games is to simply derive their normal form representations and then apply the solution concepts studied in Chapter 8. However, an important new issue arises in dynamic games: the credibility of a player's strategy. This issue is the central concern of this chapter.

Consider a vivid (although far-fetched) example: You walk into class tomorrow and your instructor, a sane but very enthusiastic game theorist, announces, "This is a very important course, and I want exclusive dedication. Anyone who does not drop every other course will be barred from the final exam and will therefore flunk." After a moment of bewilderment and some mental computation, your first thought is, "Given that I indeed prefer this course to all others, I had better follow her instructions" (after all, you have studied Chapter 8 carefully and know what a best response is). But after some further reflection, you ask yourself, "Will she really bar me from the final exam if I do not obey? This is a serious institution, and she will surely lose her job if she carries out the threat." You conclude that the answer is "no" and refuse to drop the other courses, and indeed, she ultimately does not bar you from the exam. In this example, we would say that your instructor's announced strategy, "I will bar you from the exam if you do not drop every other course," is not credible. Such empty threats are what we want to rule out as equilibrium strategies in dynamic games.

In Section 9.B, we demonstrate that the Nash equilibrium concept studied in Chapter 8 does not suffice to rule out noncredible strategies. We then introduce a stronger solution concept, known as *subgame perfect Nash equilibrium*, that helps

1. As do most parlor games.

to do so. The central idea underlying this concept is the *principle of sequential rationality*: equilibrium strategies should specify optimal behavior from any point in the game onward, a principle that is intimately related to the procedure of *backward induction.*

In Section 9.C, we show that the concept of subgame perfection is not strong enough to fully capture the idea of sequential rationality in games of imperfect information. We then introduce the notion of a *weak perfect Bayesian equilibrium* (also known as a *weak sequential equilibrium*) to push the analysis further. The central feature of a weak perfect Bayesian equilibrium is its explicit introduction of a player's *beliefs* about what may have transpired prior to her move as a means of testing the sequential rationality of the player's strategy. The modifier *weak* refers to the fact that the weak perfect Bayesian equilibrium concept imposes a *minimal* set of consistency restrictions on players' beliefs. Because the weak perfect Bayesian equilibrium concept can be too weak, we also examine some related equilibrium notions that impose stronger consistency restrictions on beliefs, discussing briefly stronger notions of *perfect Bayesian equilibrium* and, in somewhat greater detail, the concept of *sequential equilibrium.*

In Section 9.D, we go yet further by asking whether certain beliefs can be regarded as "unreasonable" in some situations, thereby allowing us to further refine our predictions. This leads us to consider the notion of *forward induction.*

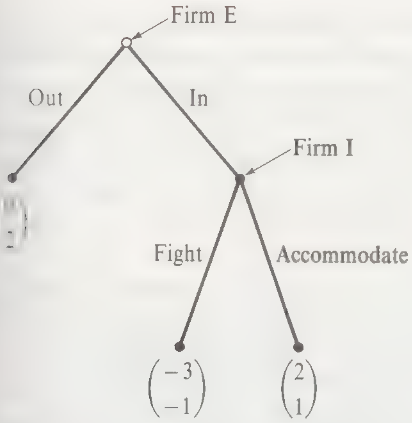
Appendix A studies finite and infinite horizon models of bilateral bargaining as an illustration of the use of subgame perfect Nash equilibrium in an important economic application. Appendix B extends the discussion in Section 9.C by examining the notion of an *extensive form trembling-hand perfect Nash equilibrium.*

We should note that—following most of the literature on this subject—all the analysis in this chapter consists of attempts to "refine" the concept of Nash equilibrium; that is, we take the position that we want our prediction to be a Nash equilibrium, and we then propose additional conditions for such an equilibrium to be a "satisfactory" prediction. However, the issues that we discuss here are not confined to this approach. We might, for example, be concerned about noncredible strategies even if we were unwilling to impose the mutually correct expectations condition of Nash equilibrium and wanted to focus instead only on rationalizable outcomes. See Bernheim (1984) and, especially, Pearce (1984) for a discussion of nonequilibrium approaches to these issues.

9.B Sequential Rationality, Backward Induction, and Subgame Perfection

We begin with an example to illustrate that in dynamic games the Nash equilibrium concept may not give sensible predictions. This observation leads us to develop a strengthening of the Nash equilibrium concept known as *subgame perfect Nash equilibrium.*

Example 9.B.1: Consider the following *predation* game. Firm E (for entrant) is considering entering a market that currently has a single incumbent (firm I). If it does so (playing "in"), the incumbent can respond in one of two ways: It can either accommodate the entrant, giving up some of its sales but causing no change in



		Firm I: Fight if Firm E Plays "In"	Firm I: Accommodate if Firm E Plays "In"
Firm E	Out	0, 2	0, 2
	In	−3, −1	2, 1

Figure 9.B.1 Extensive and normal forms for Example 9.B.1. The Nash equilibrium $(\sigma_E, \sigma_I) =$ (out, fight if firm E plays "in") involves a noncredible threat.

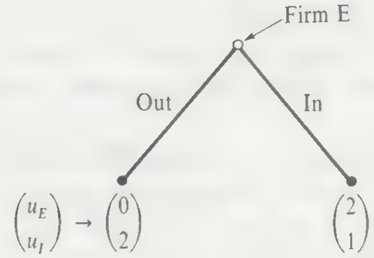
...et price, or it can fight the entrant, engaging in a costly war of predation ...atically lowers the market price. The extensive and normal form representa... of this game are depicted in Figure 9.B.1.

...ning the normal form, we see that this game has two pure strategy Nash ...ia: $(\sigma_E, \sigma_I) =$ (out, fight if firm E plays "in") and $(\sigma_E, \sigma_I) =$ (in, accommodate ... E plays "in"). Consider the first of these strategy profiles. Firm E prefers to ... the market if firm I will fight after it enters. On the other hand, "fight if ... E plays 'in'" is an optimal choice for the incumbent if firm E is playing "out." ... arguments show that the second pair of strategies is also a Nash equilibrium. ... we claim that (out, fight if firm E plays "in") is not a sensible prediction for ...game. As in the example of your instructor that we posed in Section 9.A, firm ... foresee that if it does enter, the incumbent will, in fact, find it optimal to ...odate (by doing so, firm I earns 1 rather than −1). Hence, the incumbent's ... "fight if firm E plays 'in'" is not credible. ■

...ple 9.B.1 illustrates a problem with the Nash equilibrium concept in dynamic ... In this example, the concept is, in effect, permitting the incumbent to make ...ty threat that the entrant nevertheless takes seriously when choosing its ... The problem with the Nash equilibrium concept here arises from the fact ... when the entrant plays "out," actions at decision nodes that are unreached by ... the equilibrium strategies (here, firm I's action at the decision node following ... unchosen move "in") do not affect firm I's payoff. As a result, firm I can plan ... *absolutely anything* at this decision node: Given firm E's strategy of choosing ... firm I's payoff is still maximized. *But*—and here is the crux of the matter—what ... I's strategy says it will do at the unreached node can actually *insure* that firm ...ng firm I's strategy as given, wants to play "out."

To rule out predictions such as (out, fight if firm E plays "in"), we want to insist ... players' equilibrium strategies satisfy what might be called the *principle of ... rationality*: A player's strategy should specify optimal actions *at every point ... game tree.* That is, given that a player finds herself at some point in the tree, ...ategy should prescribe play that is optimal from that point on given her ...nts' strategies. Clearly, firm I's strategy "fight if firm E plays 'in'" does not: ...ntry, the only optimal strategy for firm I is "accommodate."

... Example 9.B.1, there is a simple procedure that can be used to identify the



desirable (i.e., sequentially rational) Nash equilibrium (σ_E, σ_I) = (in, accommodate if firm E plays "in"). We first determine optimal behavior for firm I in the post-entry stage of the game; this is "accommodate." Once we have done this, we then determine firm E's optimal behavior earlier in the game given the anticipation of what will happen after entry. Note that this second step can be accomplished by considering a *reduced* extensive form game in which firm I's post-entry decision is replaced by the payoffs that will result from firm I's optimal post-entry behavior. See Figure 9.B.2. This reduced game is a very simple single-player decision problem in which firm E's optimal decision is to play "in." In this manner, we identify the sequentially rational Nash equilibrium strategy profile (σ_E, σ_I) = (in, accommodate if firm E plays "in").

This type of procedure, which involves solving first for optimal behavior at the "end" of the game (here, at the post-entry decision node) and then determining what optimal behavior is earlier in the game given the anticipation of this later behavior, is known as *backward induction* (or *backward programming*). It is a procedure that is intimately linked to the idea of sequential rationality because it insures that players' strategies specify optimal behavior at every decision node of the game.

The game in Example 9.B.1 is a member of a general class of games in which the backward induction procedure can be applied to capture the idea of sequential rationality with great generality and power: *finite games of perfect information.* These are games in which every information set contains a single decision node and there is a finite number of such nodes (see Chapter 7).[2] Before introducing a formal equilibrium concept, we first discuss the general application of the backward induction procedure to this class of games.

Backward Induction in Finite Games of Perfect Information

To apply the idea of backward induction in finite games of perfect information, we start by determining the optimal actions for moves at the final decision nodes in the tree (those for which the only successor nodes are terminal nodes). Just as in firm I's post-entry decision in Example 9.B.1, play at these nodes involves no further strategic interactions among the players, and so the determination of optimal behavior at these decision nodes involves a simple single-person decision problem. Then, given that these will be the actions taken at the final decision nodes, we can proceed to the next-to-last decision nodes and determine the optimal actions to be

2. The assumption of finiteness is important for some aspects of this analysis. We discuss this point further toward the end of the section.

...en there by players who correctly anticipate the actions that will follow at the final ...ision nodes, and so on backward through the game tree.

This procedure is readily implemented using reduced games. At each stage, after ...ing for the optimal actions at the current final decision nodes, we can derive a ... reduced game by deleting the part of the game following these nodes and ...igning to these nodes the payoffs that result from the already determined ...tinuation play.

...ample 9.B.2: Consider the three-player finite game of perfect information depicted ... Figure 9.B.3(a). The arrows in Figure 9.B.3(a) indicate the optimal play at the ... decision nodes of the game. Figure 9.B.3(b) is the reduced game formed by ...acing these final decision nodes by the payoffs that result from optimal play once ... nodes have been reached. Figure 9.B.3(c) represents the reduced game derived

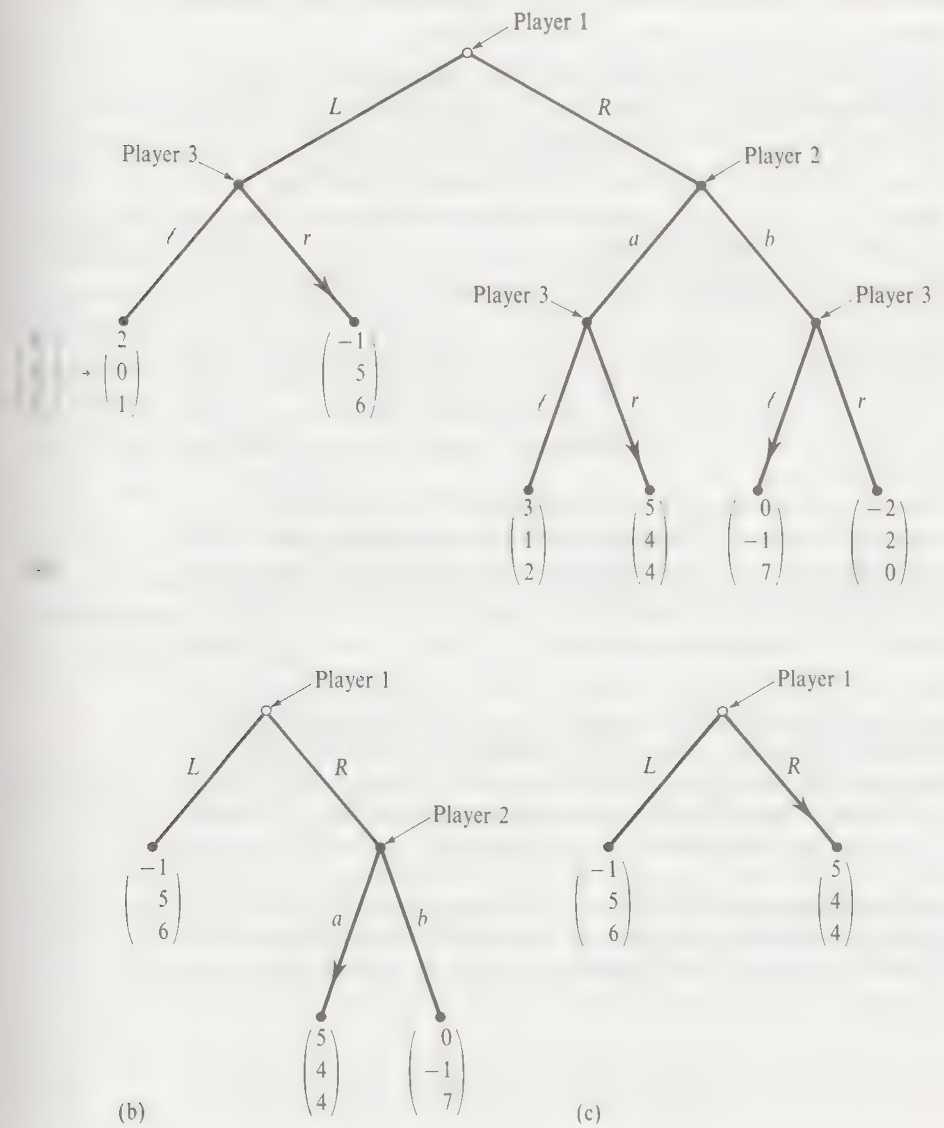


Figure 9.B.3
Reduced games in a backward induction procedure for a finite game of perfect information.
(a) Original game.
(b) First reduced game. (c) Second reduced game.

in the next stage of the backward induction procedure, when the final decision nodes of the reduced game in Figure 9.B.3(b) are replaced by the payoffs arising from optimal play at these nodes (again indicated by arrows). The backward induction procedure therefore identifies the strategy profile $(\sigma_1, \sigma_2, \sigma_3)$ in which $\sigma_1 = R$, $\sigma_2 =$ "a if player 1 plays R," and

$$\sigma_3 = \begin{cases} r \text{ if player 1 plays } L \\ r \text{ if player 1 plays } R \text{ and player 2 plays } a \\ \ell \text{ if player 1 plays } R \text{ and player 2 plays } b. \end{cases}$$

Note that this strategy profile is a Nash equilibrium of this three-player game but that the game also has other pure strategy Nash equilibria. (Exercise 9.B.3 asks you to verify these two points and to argue that these other Nash equilibria do not satisfy the principle of sequential rationality.) ■

In fact, for finite games of perfect information, we have the general result presented in Proposition 9.B.1.

Proposition 9.B.1: (*Zermelo's Theorem*) Every finite game of perfect information Γ_E has a pure strategy Nash equilibrium that can be derived through backward induction. Moreover, if no player has the same payoffs at any two terminal nodes, then there is a unique Nash equilibrium that can be derived in this manner.

Proof: First, note that in finite games of perfect information, the backward induction procedure is well defined: The player who moves at each decision node has a finite number of possible choices, so optimal actions necessarily exist at each stage of the procedure (if a player is indifferent, we can choose any of her optimal actions). Moreover, the procedure fully specifies all of the players' strategies after a finite number of stages. Second, note that if no player has the same payoffs at any two terminal nodes, then the optimal actions must be *unique* at every stage of the procedure, and so in this case the backward induction procedure identifies a unique strategy profile for the game.

What remains is to show that a strategy profile identified in this way, say $\sigma = (\sigma_1, \ldots, \sigma_I)$, is necessarily a Nash equilibrium of Γ_E. Suppose that it is not. Then there is some player i who has a deviation, say to strategy $\hat{\sigma}_i$, that strictly increases her payoff given that the other players continue to play strategies σ_{-i}. That is, letting $u_i(\sigma_i, \sigma_{-i})$ be player i's payoff function,[3]

$$u_i(\hat{\sigma}_i, \sigma_{-i}) > u_i(\sigma_i, \sigma_{-i}). \tag{9.B.1}$$

We argue that this cannot be. The proof is inductive. We shall say that decision node x has *distance* n if, among the various paths that continue from it to the terminal nodes, the maximal number of decision nodes lying between it and a terminal node is n. We let N denote the maximum distance of any decision node in the game; since Γ_E is a finite game, N is a finite number. Define $\hat{\sigma}_i(n)$ to be the strategy that plays in accordance with strategy σ_i at all nodes with distances $0, \ldots, n$, and plays in accordance with strategy $\hat{\sigma}_i$ at all nodes with distances greater than n.

By the construction of σ through the backward induction procedure, $u_i(\hat{\sigma}_i(0), \sigma_{-i}) \geq u_i(\hat{\sigma}_i, \sigma_{-i})$. That is, player i can do at least as well as she does with strategy $\hat{\sigma}_i$ by instead playing the moves specified in strategy σ_i at all nodes with distance 0 (i.e., at the final decision nodes in the game) and following strategy $\hat{\sigma}_i$ elsewhere.

3. To be precise, $u_i(\cdot)$ is player i's payoff function in the normal form derived from extensive form game Γ_E.

We now argue that if $u_i(\hat{\sigma}_i(n-1), \sigma_{-i}) \geq u_i(\hat{\sigma}_i, \sigma_{-i})$, then $u_i(\hat{\sigma}_i(n), \sigma_{-i}) \geq u_i(\hat{\sigma}_i, \sigma_{-i})$. This is straightforward. The only difference between strategy $\hat{\sigma}_i(n)$ and strategy $\hat{\sigma}_i(n-1)$ is in player i's moves at nodes with distance n. In both strategies, player i plays according to σ_i at all decision nodes that follow the distance-n nodes and in accordance with strategy $\hat{\sigma}_i$ before them. But given that all players are playing in accordance with strategy profile σ after the distance-n nodes, the moves derived for the distance-n decision nodes through backward induction, which are those in σ_i, must be optimal choices for player i at these nodes. Hence, $u_i(\hat{\sigma}_i(n), \sigma_{-i}) \geq u_i(\hat{\sigma}_i(n-1), \sigma_{-i})$.

Applying induction, we therefore have $u_i(\hat{\sigma}_i(N), \sigma_{-i}) \geq u_i(\hat{\sigma}_i, \sigma_{-i})$. But $\hat{\sigma}_i(N) = \sigma_i$, and so we have a contradiction to (9.B.1). Strategy profile σ must therefore constitute a Nash equilibrium of Γ_E. ■

Note, incidentally, that Proposition 9.B.1 establishes the existence of a pure strategy Nash equilibrium in all finite games of perfect information.

Subgame Perfect Nash Equilibria

It is clear enough how to apply the principle of sequential rationality in Example 9.B.1 and, more generally, in finite games of perfect information. Before distilling a general solution concept, however, it is useful to discuss another example. This example suggests how we might identify Nash equilibria that satisfy the principle of sequential rationality in more general games involving imperfect information.

Example 9.B.3: We consider the same situation as in Example 9.B.1 except that firms E and I now play a simultaneous-move game after entry, each choosing either "fight" or "accommodate." The extensive and normal form representations are depicted in Figure 9.B.4.

Examining the normal form, we see that in this game there are three pure strategy Nash equilibria (σ_E, σ_I):[4]

((out, accommodate if in), (fight if firm E plays "in")),

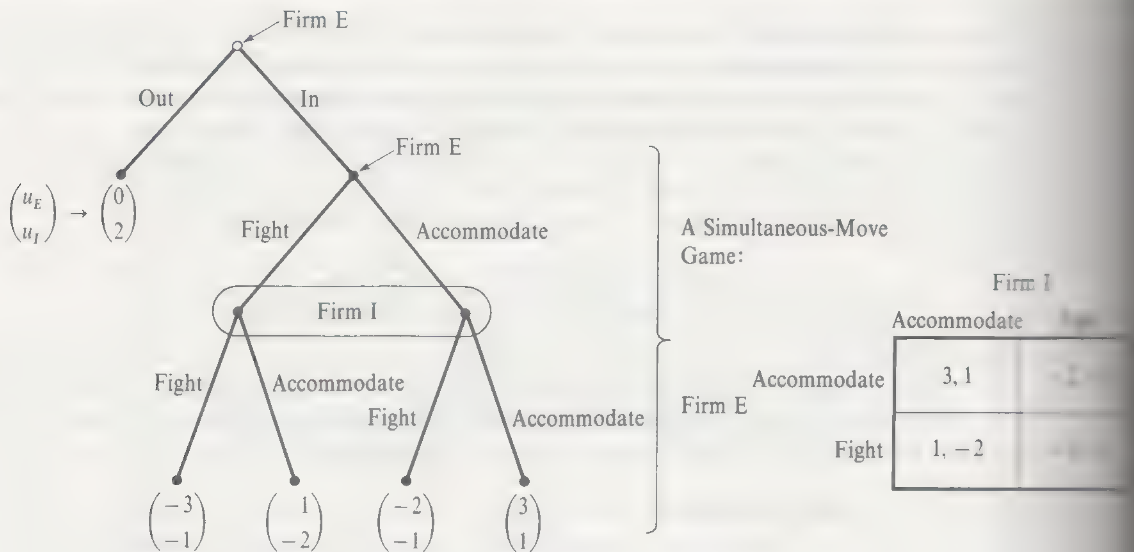
((out, fight if in), (fight if firm E plays "in")),

((in, accommodate if in), (accommodate if firm E plays "in")).

Note, however, that (accommodate, accommodate) is the sole Nash equilibrium in the simultaneous-move game that follows entry. Thus, the firms should expect that they will both play "accommodate" following firm E's entry.[5] But if this is so, firm E

4. The entrant's strategy in the first two equilibria may appear odd. Firm E is planning to take an action conditional on entering while at the same time planning not to enter. Recall from Section 7.D, however, that a strategy is a *complete contingent plan*. Indeed, the reason we have insisted on this requirement is precisely the need to test the sequential rationality of a player's strategy.

5. Recall that throughout this chapter we maintain the assumption that rational players always play a Nash equilibrium in any strategic situation in which they find themselves (i.e., we assume that players will have mutually correct expectations). Two points about this assumption are worth noting. First, some justifications for a Nash equilibrium may be less compelling in the context of dynamic games. For example, if players never reach certain parts of a game, the stable social convention argument given in Section 8.D may no longer provide a good reason for believing that a Nash equilibrium would be played *if* that part of the game tree were reached. Second, the idea of sequential rationality can still have force even if we do not make this assumption. For example, here we would reach the same conclusion even if we assumed only that neither player would play an iteratively strictly dominated strategy in the post-entry simultaneous-move game.



		Firm I	
		Accommodate if E Plays "In"	Fight if E Plays "In"
Firm E	Out, Accommodate if In	0, 2	0, 2
	Out, Fight if In	0, 2	0, 2
	In, Accommodate if In	3, 1	−2, −1
	In, Fight if In	1, −2	−3, −1

Figure 9.B.4 Extensive and normal forms for Example 9.B.3. A sequentially rational Nash equilibrium firms play "accommodate" after entry.

should enter. The logic of sequential rationality therefore suggests that only the last of the three equilibria is a reasonable prediction in this game. ■

The requirement of sequential rationality illustrated in this and the preceding examples is captured by the notion of a *subgame perfect Nash equilibrium* [introduced by Selten (1965)]. Before formally defining this concept, however, we need to specify what a *subgame* is.

Definition 9.B.1: A *subgame* of an extensive form game Γ_E is a subset of the game having the following properties:

(i) It begins with an information set containing a single decision node, contains all the decision nodes that are successors (both immediate and later) of this node, and contains *only* these nodes.
(ii) If decision node x is in the subgame, then every $x' \in H(x)$ is also, where $H(x)$ is the information set that contains decision node x. (That is, there are no "broken" information sets.)

Note that according to Definition 9.B.1, the game as a whole is a subgame, as

some strict subsets of the game.[6] For example, in Figure 9.B.1, there are two ...s: the game as a whole and the part of the game tree that begins with and ... firm I's decision node. The game in Figure 9.B.4 also has two subgames: the ... whole and the part of the game beginning with firm E's post-entry decision ... In Figure 9.B.5, the dotted lines indicate three parts of the game of Figure ... that are *not* subgames.

... note that in a finite game of perfect information, every decision node ...s a subgame. (Exercise 9.B.1 asks you to verify this fact for the game of ... 9.B.2.)

... key feature of a subgame is that, contemplated in isolation, it is a game in ... right. We can therefore apply to it the idea of Nash equilibrium predictions. ... discussion that follows, we say that a strategy profile σ in extensive form game ...s a Nash equilibrium in a particular subgame of Γ_E if the moves specified ... information sets within the subgame constitute a Nash equilibrium when ... subgame is considered in isolation.

... 9.B.2: A profile of strategies $\sigma = (\sigma_1, \ldots, \sigma_I)$ in an I-player extensive form ... s a *subgame perfect Nash equilibrium* (SPNE) if it induces a Nash ...rium in every subgame of Γ_E.

Note that any SPNE is a Nash equilibrium (since the game as a whole is a ...) but that not every Nash equilibrium is subgame perfect.

... 9.B.2: Consider a game Γ_E in extensive form. Argue that:

(a) If the only subgame is the game as a whole, then every Nash equilibrium is subgame perfect.

(b) A subgame perfect Nash equilibrium induces a subgame perfect Nash equilibrium in every subgame of Γ_E.

The SPNE concept isolates the reasonable Nash equilibria in Examples 9.B.1 and ... In Example 9.B.1, any subgame perfect Nash equilibrium must have firm I ... "accommodate if firm E plays 'in'" because this is firm I's strictly dominant ... in the subgame following entry. Likewise, in Example 9.B.3, any SPNE must ... the firms both playing "accommodate" after entry because this is the unique ... equilibrium in this subgame.

Note also that in finite games of perfect information, such as the games of ... 9.B.1 and 9.B.2, the set of SPNEs coincides with the set of Nash equilibria ... be derived through the backward induction procedure. Recall, in particular, ... finite games of perfect information every decision node initiates a subgame. ... in any SPNE, the strategies must specify actions at each of the final decision ... the game that are optimal in the single-player subgame that begins there. ... that this must be the play at the final decision nodes in any SPNE, consider ... the subgames starting at the next-to-last decision nodes. Nash equilibrium ... in these subgames, which is required in any SPNE, must have the players who

6. In the literature, the term *proper subgame* is sometimes used with the same meaning we assign ... We choose to use the unqualified term *subgame* here to make clear that the game itself ...

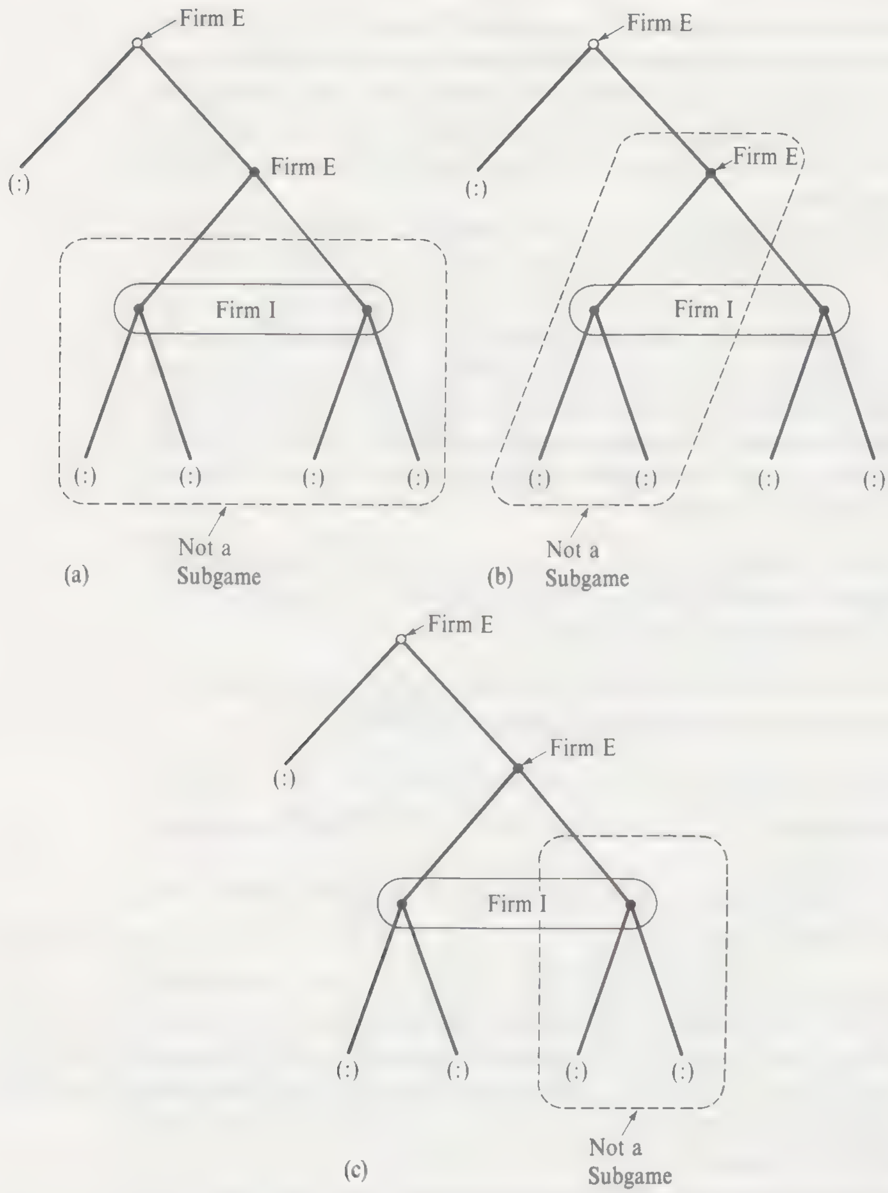


Figure 9.B.5
Three par[illegible]
game in Fig[illegible]
that are not [illegible]

move at these next-to-last nodes choosing optimal strategies given the play that will occur at the last nodes. And so on. An implication of this fact and Proposition 9.B.1 is therefore the result stated in Proposition 9.B.2.

Proposition 9.B.2: Every finite game of perfect information Γ_E has a pure strategy subgame perfect Nash equilibrium. Moreover, if no player has the same payoffs at any two terminal nodes, then there is a unique subgame perfect Nash equilibrium.[7]

7. The result can also be seen directly from Proposition 9.B.1. Just as the strategy profile derived using the backward induction procedure constitutes a Nash equilibrium in the game as a whole, it is also a Nash equilibrium in every subgame.

In fact, to identify the set of subgame perfect Nash equilibria in a general (finite) dynamic game Γ_E, we can use a generalization of the backward induction procedure. This *generalized backward induction procedure* works as follows:

1. Start at the end of the game tree, and identify the Nash equilibria for each of the *final* subgames (i.e., those that have no other subgames nested within them).
2. Select one Nash equilibrium in each of these final subgames, and derive the reduced extensive form game in which these final subgames are replaced by the payoffs that result in these subgames when players use these equilibrium strategies.
3. Repeat steps 1 and 2 for the reduced game. Continue the procedure until every move in Γ_E is determined. This collection of moves at the various information sets of Γ_E constitutes a profile of SPNE strategies.
4. If multiple equilibria are never encountered in any step of this process, this profile of strategies is the unique SPNE. If multiple equilibria are encountered, the full set of SPNEs is identified by repeating the procedure for each possible equilibrium that could occur for the subgames in question.

The formal justification for using this generalized backward induction procedure to identify the set of SPNEs comes from the result shown in Proposition 9.B.3.

Proposition 9.B.3: Consider an extensive form game Γ_E and some subgame S of Γ_E. Suppose that strategy profile σ^S is an SPNE in subgame S, and let $\hat{\Gamma}_E$ be the reduced game formed by replacing subgame S by a terminal node with payoffs equal to those arising from play of σ^S. Then:

(i) In any SPNE σ of Γ_E in which σ^S is the play in subgame S, players' moves at information sets outside subgame S must constitute an SPNE of reduced game $\hat{\Gamma}_E$.

(ii) If $\hat{\sigma}$ is an SPNE of $\hat{\Gamma}_E$, then the strategy profile σ that specifies the moves in σ^S at information sets in subgame S and that specifies the moves in $\hat{\sigma}$ at information sets not in S is an SPNE of Γ_E.

Proof: (i) Suppose that strategy profile σ specifies play at information sets outside subgame S that does not constitute an SPNE of reduced game $\hat{\Gamma}_E$. Then there exists a subgame of $\hat{\Gamma}_E$ in which σ does not induce a Nash equilibrium. In this subgame of $\hat{\Gamma}_E$, some player has a deviation that improves her payoff, taking as given the strategies of her opponents. But then it must be that this player also has a profitable deviation in the corresponding subgame of game Γ_E. She makes the same alterations in her moves at information sets not in S and leaves her moves at information sets in S unchanged. Hence, σ could not be an SPNE of the overall game Γ_E.

(ii) Suppose that $\hat{\sigma}$ is an SPNE of reduced game $\hat{\Gamma}_E$, and let σ be the strategy in the overall game Γ_E formed by specifying the moves in σ^S at information sets in subgame S and the moves in $\hat{\sigma}$ at information sets not in S. We argue that σ induces a Nash equilibrium in every subgame of Γ_E. This follows immediately from the construction of σ for subgames of Γ_E that either lie entirely in subgame S or never intersect with subgame S (i.e., that do not have subgame S nested within them). So consider any subgame that has subgame S nested within it. If some player i has a profitable deviation in this subgame given her opponent's strategies, then she must also have a profitable deviation that leaves her moves within subgame S unchanged because, by hypothesis, a player does best within subgame S by playing the moves specified by strategy profile σ^S given that her opponents do so. But if she has such a profitable deviation,

then she must have a profitable deviation in the corresponding subgame of reduced game $\hat{\Gamma}_E$, in contradiction to $\hat{\sigma}$ being an SPNE of $\hat{\Gamma}_E$. ■

Note that for the final subgames of Γ_E, the set of Nash equilibria and SPNEs coincide, because these subgames contain no nested subgames. Identifying Nash equilibria in these final subgames therefore allows us to begin the inductive application of Proposition 9.B.3.

This generalized backward induction procedure reduces to our previous backward induction procedure in the case of games of perfect information. But it also applies to games of imperfect information. Example 9.B.3 provides a simple illustration. There we can identify the unique SPNE by first identifying the unique Nash equilibrium in the post-entry subgame: (accommodate, accommodate). Having done this, we can replace this subgame with the payoffs that result from equilibrium play in it. The reduced game that results is then much the same as that shown in Figure 9.B.2, the only difference being that firm E's payoff from playing "in" is now 3 instead of 2. Hence, in this manner, we can derive the unique SPNE of Example 9.B.3: $(\sigma_E, \sigma_I) =$ ((in, accommodate if in), (accommodate if firm E plays "in")).

The game in Example 9.B.3 is simple to solve in two respects. First, there is a unique equilibrium in the post-entry subgame. If this were not so, behavior earlier in the game could depend on *which* equilibrium resulted after entry. Example 9.B.4 illustrates this point:[8]

Example 9.B.4: *The Niche Choice Game.* Consider a modification of Example 9.B.3 in which instead of having the two firms choose whether to fight or accommodate each other, we suppose that there are actually two niches in the market, one large and one small. After entry, the two firms decide simultaneously which niche they will be in. For example, the niches might correspond to two types of customers, and the firms may be deciding to which type they are targeting their product design. Both firms lose money if they choose the same niche, with more lost if it is the small niche. If they choose different niches, the firm that targets the large niche earns a profit, and the firm with the small niche incurs a loss, but a smaller loss than if the two firms targeted the same niche. The extensive form of this game is depicted in Figure 9.B.6.

To determine the SPNE of this game, consider the post-entry subgame first. There are two pure strategy Nash equilibria of this simultaneous-move game: (large niche, small niche) and (small niche, large niche).[9] In any pure strategy SPNE, the firms' strategies must induce one of these two Nash equilibria in the post-entry subgame. Suppose, first, that the firms will play (large niche, small niche). In this case, the payoffs from reaching the post-entry subgame are $(u_E, u_I) = (1, -1)$, and the reduced game is as depicted in Figure 9.B.7(a). The entrant optimally chooses to enter in this

---

8. Similar issues can arise in games of perfect information when a player is indifferent between two actions. However, the presence of multiple equilibria in subgames involving simultaneous play is, in a sense, a more robust phenomenon. Multiple equilibria are generally robust to small changes in players' payoffs, but ties in games of perfect information are not.

9. We restrict attention here to pure strategy SPNEs. There is also a mixed strategy Nash equilibrium in the post-entry subgame. Exercise 9.B.6 asks you to investigate the implications of this mixed strategy play being the post-entry equilibrium behavior.

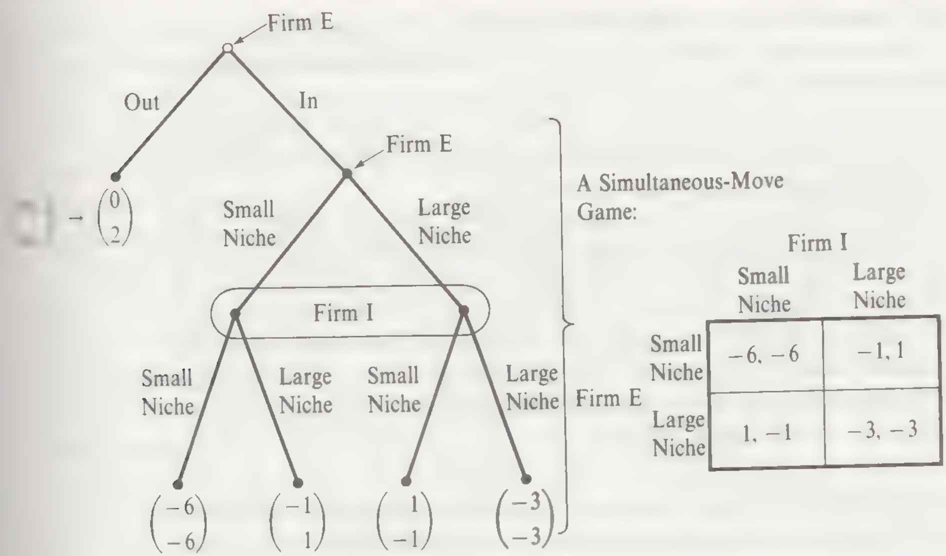


Figure 9.B.6 Extensive form for the Niche Choice game. The post-entry subgame has multiple Nash equilibria.

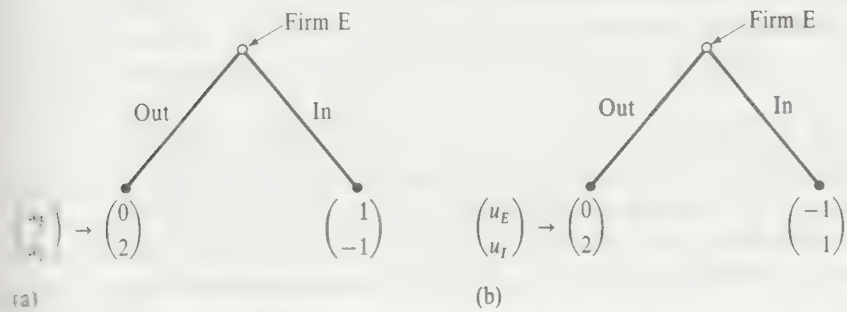


Figure 9.B.7 Reduced games after identifying (pure strategy) Nash equilibria in the post-entry subgame of the Niche Choice game. (a) Reduced game if (large niche, small niche) is post-entry equilibrium. (b) Reduced game if (small niche, large niche) is post-entry equilibrium.

... Hence, one SPNE is $(\sigma_E, \sigma_I) =$ ((in, large niche if in), (small niche if firm E plays "in")).

Now suppose that the post-entry play is (small niche, large niche). Then the payoffs from reaching the post-entry game are $(u_E, u_I) = (-1, 1)$, and the reduced game is that depicted in Figure 9.B.7(b). The entrant optimally chooses not to enter in this case. Hence, there is a second pure strategy SPNE: $(\sigma_E, \sigma_I) =$ ((out, small niche if in), (large niche if firm E plays "in")). ■

A second sense in which the game in Example 9.B.3 is simple to solve is that it involves only one subgame other than the game as a whole. Like games of perfect information, a game with imperfect information may in general have *many* subgames, with one subgame nested within another, and that larger subgame nested within a still larger one, and so on.

One interesting class of imperfect information games in which the generalized backward induction procedure gives a very clean conclusion is described in Proposition 9.B.4.

Proposition 9.B.4: Consider an I-player extensive form game Γ_E involving successive play of T simultaneous-move games, $\Gamma_N^t = [I, \{\Delta(S_i^t)\}, \{u_i^t(\cdot)\}]$ for $t = 1, \ldots, T$, with the players observing the pure strategies played in each game immediately after its play is concluded. Assume that each player's payoff is equal to the sum of her payoffs in the plays of the T games. If there is a unique Nash equilibrium

in each game Γ_N^t, say $\sigma^t = (\sigma_1^t, \ldots, \sigma_I^t)$, then there is a unique SPNE of Γ_E and it consists of each player i playing strategy σ_i^t in each game Γ_N^t regardless of what has happened previously.

Proof: The proof is by induction. The result is clearly true for $T = 1$. Now suppose it is true for all $T \leq n - 1$. We will show that it is true for $T = n$.

We know by hypothesis that in any SPNE of the overall game, after play of game Γ_N^1 the play in the remaining $n - 1$ simultaneous-move games must simply involve play of the Nash equilibrium of each game (since any SPNE of the overall game induces an SPNE in each of its subgames). Let player i earn G_i from this equilibrium play in these $n - 1$ games. Then in the reduced game that replaces all the subgames that follow Γ_N^1 with their equilibrium payoffs, player i earns an overall payoff of $u_i(s_1^1, \ldots, s_I^1) + G_i$ if $(s_1^1, \ldots, s_I^1)$ is the profile of pure strategies played in game Γ_N^1. The unique Nash equilibrium of this reduced game is clearly σ^1. Hence, the result is also true for $T = n$. ■

The basic idea behind Proposition 9.B.4 is an application of backward induction logic: Play in the last game must result in the unique Nash equilibrium of that game being played because at that point players essentially face just that game. But if play in the last game is predetermined, then when players play the next-to-last game, it is again as if they were playing just *that* game in isolation (think of the case where $T = 2$). And so on.

An interesting aspect of Proposition 9.B.4 is the way the SPNE concept rules out history dependence of strategies in the class of games considered there. In general, a player's strategy could potentially promise later rewards or punishments to other players if they take particular actions early in the game. But as long as each of the component games has a unique Nash equilibrium, SPNE strategies cannot be history dependent.[10]

Exercises 9.B.9 to 9.B.11 provide some additional examples of the use of the subgame perfect Nash equilibrium concept. In Appendix A we also study an important economic application of subgame perfection to a finite game of perfect information (albeit one with an infinite number of possible moves at some decision nodes): a finite horizon model of bilateral bargaining.

Up to this point, our analysis has assumed that the game being studied is finite. This has been important because it has allowed us to identify subgame perfect Nash equilibria by starting at the end of the game and working backward. As a general matter, in games in which there can be an infinite sequence of moves (so that some paths through the tree never reach a terminal node), the definition of a subgame perfect Nash equilibrium remains that given in Definition 9.B.2: Strategies must induce a Nash equilibrium in every subgame. Nevertheless, the lack of a definite finite point of termination of the game can reduce the power of the SPNE concept because we can no longer use the end of the game to pin down behavior. In games in which there is always a future, a wide range of behaviors can sometimes be justified as sequentially rational (i.e., as part of an SPNE). A striking example of this sort arises in

10. This lack of history dependence depends importantly on the uniqueness assumption of Proposition 9.B.4. With multiple Nash equilibria in the component games, we can get outcomes that are not merely the repeated play of the static Nash equilibria. (See Exercise 9.B.9 for an example.)

Chapter 12 and its Appendix A when we consider *infinitely repeated games* in the context of studying oligopolistic pricing.

Nevertheless, it is not always the case that an infinite horizon weakens the power of the subgame perfection criterion. In Appendix A of this chapter, we study an infinite horizon model of bilateral bargaining in which the SPNE concept predicts a unique outcome, and this outcome coincides with the limiting outcome of the corresponding finite horizon bargaining model as the horizon grows long.

The methods used to identify subgame perfect Nash equilibria in infinite horizon games are varied. Sometimes, the method involves showing that the game can effectively be truncated because after a certain point it is obvious what equilibrium play must be (see Exercise 9.B.11). In other situations, the game possesses a stationarity property that can be exploited; the analysis of the infinite horizon bilateral bargaining model in Appendix A is one example of this kind.

After the preceding analysis, the logic of sequential rationality may seem unassailable. But things are not quite so clear. For example, nowhere could the principle of sequential rationality seem on more secure footing than in finite games of perfect information. But chess is a game of this type (the game ends if 50 moves occur without a piece being taken or a pawn being moved), and so its "solution" should be simple to predict. Of course, it is exactly players' *inability* to do so that makes it an exciting game to play. The same could be said even of the much simpler game of Chinese checkers. It is clear that in practice, players may be only boundedly rational. As a result, we might feel more comfortable with our rationality hypotheses in games that are relatively simple, in games where repetition helps players learn to think through the game, or in games where large stakes encourage players to do so as much as possible. Of course, the possibility of bounded rationality is not a concern limited to dynamic games and subgame perfect Nash equilibria; it is also relevant for simultaneous-move games containing many possible strategies.

There is, however, an interesting tension present in the SPNE concept that is related to this bounded rationality issue and that does not arise in the context of simultaneous-move games. In particular, the SPNE concept insists that players should play an SPNE wherever they find themselves in the game tree, even after a sequence of events that is contrary to the predictions of the theory. To see this point starkly, consider the following example due to Rosenthal (1981), known as the *Centipede game.*

Example 9.B.5: *The Centipede Game.* In this finite game of perfect information, there are two players, 1 and 2. The players each start with 1 dollar in front of them. They alternate saying "stop" or "continue," starting with player 1. When a player says "continue," 1 dollar is taken by a referee from her pile and 2 dollars are put in her opponent's pile. As soon as either player says "stop," play is terminated, and each player receives the money currently in her pile. Alternatively, play stops if both players' piles reach 100 dollars. The extensive form for this game is depicted in Figure 9.B.8.

The unique SPNE in this game has both players saying "stop" whenever it is their turn, and the players each receive 1 dollar in this equilibrium. To see this, consider player 2's move at the final decision node of the game (after the players have said "continue" a total of 197 times). Her optimal move if play reaches this point is to say "stop"; by doing so, she receives 101 dollars compared with a payoff of 100 dollars if she says "continue." Now consider what happens if play reaches the next-to-last decision node. Player 1, anticipating player 2's move at the final decision node, also says "stop"; doing so, she earns 99 dollars, compared with 98 dollars if she says "continue." Continuing backward through the tree in this fashion, we identify saying "stop" as the optimal move at every decision node.

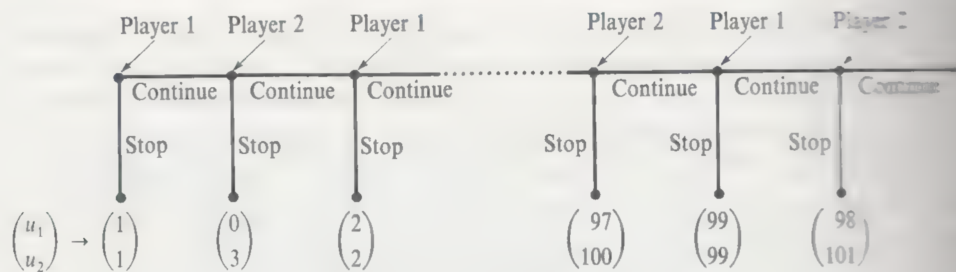


Figure 9.B.8 The Centipede game.

A striking aspect of the SPNE in the Centipede game is how bad it is for the players. They each get 1 dollar, whereas they might get 100 dollars by repeatedly saying "continue."

Is this (unique) SPNE in the Centipede game a reasonable prediction? Consider player 1's initial decision to say "stop." For this to be rational, player 1 must be pretty sure that if instead she says "continue," player 2 will say "stop" at her first turn. Indeed, "continue" would be better for player 1 as long as she could be sure that player 2 would say "continue" at her next move. Why might player 2 respond to player 1 saying "continue" by also saying "continue"? First, as we have pointed out, player 2 might not be fully rational, and so she might not have done the backward induction computation assumed in the SPNE concept. More interestingly, however, once she sees that player 1 has chosen "continue"—an event that should never happen according to the SPNE prediction—she might entertain the possibility that player 1 is not rational in the sense demanded by the SPNE concept. If, as a result, she thinks that player 1 would say "continue" at her next move if given the chance, then player 2 would want to say "continue" herself. The SPNE concept denies this possibility, instead assuming that at any point in the game, players will assume that the remaining play of the game will be an SPNE even if play up to that point has contradicted the theory. One way of resolving this tension is to view the SPNE theory as treating any deviation from prescribed play as the result of an extremely unlikely "mistake" that is unlikely to occur again. In Appendix B, we discuss one concept that makes this idea explicit. ■

9.C Beliefs and Sequential Rationality

Although subgame perfection is often very useful in capturing the principle of sequential rationality, sometimes it is not enough. Consider Example 9.C.1's adaptation of the entry game studied in Example 9.B.1.

Example 9.C.1: We now suppose that there are two strategies firm E can use to enter, "in_1" and "in_2," and that the incumbent is unable to tell which strategy firm E has used if entry occurs. Figure 9.C.1 depicts this game and its payoffs.

As in the original entry game in Example 9.B.1, there are two pure strategy Nash equilibria here: (out, fight if entry occurs) and (in_1, accommodate if entry occurs). Once again, however, the first of these does not seem very reasonable; regardless of what entry strategy firm E has used, the incumbent prefers to accommodate once entry has occurred. *But the criterion of subgame perfection is of absolutely no use here*: Because the only subgame is the game as a whole, both pure strategy Nash equilibria are subgame perfect. ■

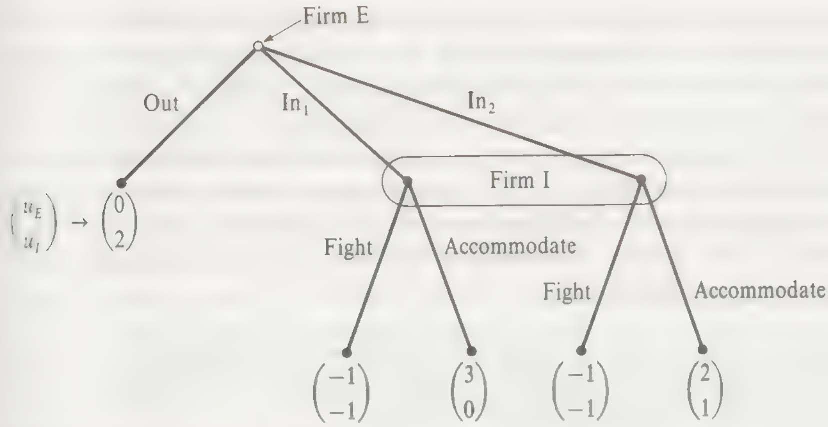


Figure 9.C.1
Extensive form for Example 9.C.1. The SPNE concept may fail to insure sequential rationality.

How can we eliminate the unreasonable equilibrium here? One possibility, which is in the spirit of the principle of sequential rationality, might be to insist that the incumbent's action after entry be optimal for *some belief* that she might have about which entry strategy was used by firm E. Indeed, in Example 9.C.1, "fight if entry occurs" is not an optimal choice for *any* belief that firm I might have. This suggests that we may be able to make some progress by formally considering players' beliefs and using them to test the sequential rationality of players' strategies.

We now introduce a solution concept, which we call a *weak perfect Bayesian equilibrium* [Myerson (1991) refers to this same concept as a *weak sequential equilibrium*], that extends the principle of sequential rationality by formally introducing the notion of beliefs.[11] It requires, roughly, that at any point in the game, a player's strategy prescribe optimal actions from that point on given her opponents' strategies and her beliefs about what has happened so far in the game and that her beliefs be consistent with the strategies being played.

To express this notion formally, we must first formally define the two concepts that are its critical components: the notions of a *system of beliefs* and the *sequential rationality of strategies.* Beliefs are simple.

Definition 9.C.1: A *system of beliefs* μ in extensive form game Γ_E is a specification of a probability $\mu(x) \in [0, 1]$ for each decision node x in Γ_E such that

$$\textstyle\sum_{x \in H} \mu(x) = 1$$

for all information sets H.

A system of beliefs can be thought of as specifying, for each information set, a probabilistic assessment by the player who moves at that set of the relative likelihoods of being at each of the information set's various decision nodes, conditional upon play having reached that information set.

11. The concept of a *perfect Bayesian equilibrium* was first developed to capture the requirements of sequential rationality in dynamic games with incomplete information, that is (using the terminology introduced in Section 8.E), in dynamic Bayesian games. The *weak perfect Bayesian equilibrium* concept is a variant that is introduced here primarily for pedagogic purposes (the reason for the modifier *weak* will be made clear later in this section). Myerson (1991) refers to this same concept as a *weak sequential equilibrium* because it may also be considered a weak variant of the *sequential equilibrium* concept introduced in Definition 9.C.4.

To define sequential rationality, it is useful to let $E[u_i \mid H, \mu, \sigma_i, \sigma_{-i}]$ denote player i's expected utility starting at her information set H if her beliefs regarding the conditional probabilities of being at the various nodes in H are given by μ, if she follows strategy σ_i, and if her rivals use strategies σ_{-i}. [We will not write out the formula for this expression explicitly, although it is conceptually straightforward: Pretend that the probability distribution $\mu(x)$ over nodes $x \in H$ is generated by nature; then player i's expected payoff is determined by the probability distribution that is induced on the terminal nodes by the combination of this initial distribution plus the players' strategies from this point on.]

Definition 9.C.2: A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_I)$ in extensive form game Γ_E is *sequentially rational at information set H given a system of beliefs* μ if, denoting by $\iota(H)$ the player who moves at information set H, we have

$$E[u_{\iota(H)} \mid H, \mu, \sigma_{\iota(H)}, \sigma_{-\iota(H)}] \geq E[u_{\iota(H)} \mid H, \mu, \tilde{\sigma}_{\iota(H)}, \sigma_{-\iota(H)}]$$

for all $\tilde{\sigma}_{\iota(H)} \in \Delta(S_{\iota(H)})$. If strategy profile σ satisfies this condition for *all* information sets H, then we say that σ is *sequentially rational given belief system* μ.

In words, a strategy profile $\sigma = (\sigma_1, \ldots, \sigma_I)$ is sequentially rational if no player finds it worthwhile, once one of her information sets has been reached, to revise her strategy given her beliefs about what has already occurred (as embodied in μ) and her rivals' strategies.

With these two notions, we can now define a weak perfect Bayesian equilibrium. The definition involves two conditions: First, strategies must be sequentially rational given beliefs. Second, whenever possible, beliefs must be consistent with the strategies. The idea behind the consistency condition on beliefs is much the same as the idea underlying the concept of Nash equilibrium (see Section 8.D): In an equilibrium, players should have correct beliefs about their opponents' strategy choices.

To motivate the specific consistency requirement on beliefs to be made in the definition of a weak perfect Bayesian equilibrium, consider how we might define the notion of consistent beliefs in the special case in which each player's equilibrium strategy assigns a strictly positive probability to each possible action at every one of her information sets (known as a *completely mixed strategy*).[12] In this case, every information set in the game is reached with positive probability. The natural notion of beliefs being consistent with the play of the equilibrium strategy profile σ is in this case straightforward: For each node x in a given player's information set H, the player should compute the probability of reaching that node given play of strategies σ, Prob $(x \mid \sigma)$, and she should then assign conditional probabilities of being at each of these nodes given that play has reached this information set using *Bayes' rule*:[13]

$$\text{Prob}\,(x \mid H, \sigma) = \frac{\text{Prob}\,(x \mid \sigma)}{\sum_{x' \in H} \text{Prob}\,(x' \mid \sigma)}.$$

12. Equivalently, a completely mixed strategy can be thought of as a strategy that assigns a strictly positive probability to each of the player's pure strategies in the normal form derived from extensive form game Γ_E.

13. Bayes' rule is a basic principle of statistical inference. See, for example, DeGroot (1970), where it is referred to as *Bayes' theorem*.

As a concrete example, suppose that in the game in Example 9.C.1, firm E is using the completely mixed strategy that assigns a probability of $\frac{1}{4}$ to "out," $\frac{1}{2}$ to "in$_1$," and $\frac{1}{4}$ to "in$_2$." Then the probability of reaching firm I's information set given this strategy is $\frac{3}{4}$. Using Bayes' rule, the probability of being at the left node of firm I's information set conditional on this information set having been reached is $\frac{2}{3}$, and the conditional probability of being at the right node in the set is $\frac{1}{3}$. For firm I's beliefs following entry to be consistent with firm E's strategy, firm I's beliefs should assign exactly these probabilities.

The more difficult issue arises when players are not using completely mixed strategies. In this case, some information sets may no longer be reached with positive probability, and so we cannot use Bayes' rule to compute conditional probabilities for the nodes in these information sets. At an intuitive level, this problem corresponds to the idea that even if players were to play the game repeatedly, the equilibrium play would generate no experience on which they could base their beliefs at these information sets. The weak perfect Bayesian equilibrium concept takes an agnostic view toward what players should believe if play were to reach these information sets unexpectedly. In particular, it allows us to assign *any* beliefs at these information sets. It is in this sense that the modifier *weak* is appropriately attached to this concept.

We can now give a formal definition.

Definition 9.C.3: A profile of strategies and system of beliefs (σ, μ) is a *weak perfect Bayesian equilibrium* (weak PBE) in extensive form game Γ_E if it has the following properties:

(i) The strategy profile σ is sequentially rational given belief system μ.

(ii) The system of beliefs μ is derived from strategy profile σ through Bayes' rule whenever possible. That is, for any information set H such that $\text{Prob}(H \mid \sigma) > 0$ (read as "the probability of reaching information set H is positive under strategies σ"), we must have

$$\mu(x) = \frac{\text{Prob}(x \mid \sigma)}{\text{Prob}(H \mid \sigma)} \quad \text{for all } x \in H.$$

It should be noted that the definition formally incorporates beliefs as part of an equilibrium by identifying a *strategy–beliefs pair* (σ, μ) as a weak perfect Bayesian equilibrium. In the literature, however, it is not uncommon to see this treated a bit loosely: a set of strategies σ will be referred to as an equilibrium with the meaning that there is at least one associated set of beliefs μ such that (σ, μ) satisfies Definition 9.C.3. At times, however, it can be very useful to be more explicit about what these beliefs are, such as when testing them against some of the "reasonableness" criteria that we discuss in Section 9.D.

A useful way to understand the relationship between the weak PBE concept and that of Nash equilibrium comes in the characterization of Nash equilibrium given in Proposition 9.C.1.

Proposition 9.C.1: A strategy profile σ is a Nash equilibrium of extensive form game Γ_E if and only if there exists a system of beliefs μ such that

(i) The strategy profile σ is sequentially rational given belief system μ *at all information sets H such that Prob $(H \mid \sigma) > 0$.*

(ii) The system of beliefs μ is derived from strategy profile σ through Bayes rule whenever possible.

Exercise 9.C.1 asks you to prove this result. The italicized portion of condition (i) is the only change from Definition 9.C.3: For a Nash equilibrium, we require sequential rationality only on the equilibrium path. Hence, a weak perfect Bayesian equilibrium of game Γ_E is a Nash equilibrium, but not every Nash equilibrium is a weak PBE.

We now illustrate the application of the weak PBE concept in several examples. We first consider how the concept performs in Example 9.C.1.

Example 9.C.1 Continued: Clearly, firm I must play "accommodate if entry occurs" in any weak perfect Bayesian equilibrium because that is firm I's optimal action starting at its information set for *any* system of beliefs. Thus, the Nash equilibrium strategies (out, fight if entry occurs) cannot be part of any weak PBE.

What about the other pure strategy Nash equilibrium, (in_1, accommodate if entry occurs)? To show that this strategy profile *is* part of a weak PBE, we need to supplement these strategies with a system of beliefs that satisfy criterion (ii) of Definition 9.C.3 and that lead these strategies to be sequentially rational. Note first that to satisfy criterion (ii), the incumbent's beliefs must assign probability 1 to being at the left node in her information set because this information set is reached with positive probability given the strategies (in_1, accommodate if entry occurs) [a specification of beliefs at this information set fully describes a system of beliefs in this game because the only other information set is a singleton]. Moreover, these strategies are, indeed, sequentially rational given this system of beliefs. In fact, this strategy beliefs pair is the unique weak PBE in this game (pure or mixed). ■

Examples 9.C.2 and 9.C.3 provide further illustrations of the application of the weak PBE concept.

Example 9.C.2: Consider the following "joint venture" entry game: Now there is a second potential entrant E2. The story is as follows: Firm E1 has the essential capability to enter the market but lacks some important capability that firm E2 has. As a result, E1 is considering proposing a joint venture with E2 in which E2 shares its capability with E1 and the two firms split the profits from entry. Firm E1 has three initial choices: enter directly on its own, propose a joint venture with E2, or stay out of the market. If it proposes a joint venture, firm E2 can either accept or decline. If E2 accepts, then E1 enters with E2's assistance. If not, then E1 must decide whether to enter on its own. The incumbent can observe whether E1 has entered, but not whether it is with E2's assistance. Fighting is the best response for the incumbent if E1 is unassisted (E1 can then be wiped out quickly) but is not optimal for the incumbent if E1 is assisted (E1 is then a tougher competitor). Finally, if E1 is unassisted, it wants to enter only if the incumbent accommodates; but if E1 is assisted by E2, then because it will be such a strong competitor, its entry is profitable regardless of whether the incumbent fights. The extensive form of this game is depicted in Figure 9.C.2.

To identify the weak PBE of this game note first that, in any weak PBE, firm E2 must accept the joint venture if firm E1 proposes it because E2 is thereby assured of a positive payoff regardless of firm I's strategy. But if so, then in any weak PBE

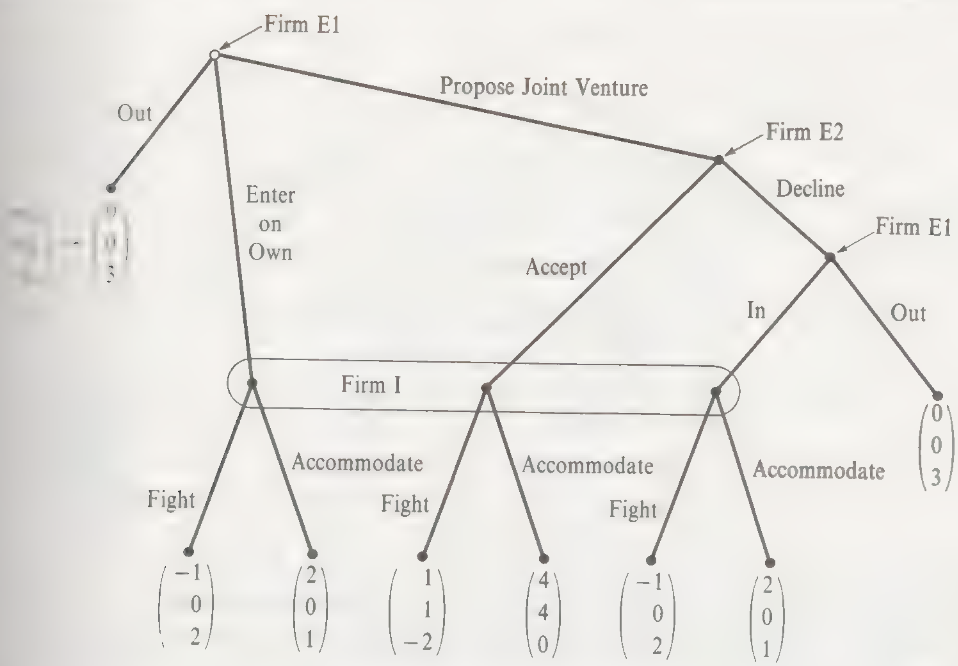


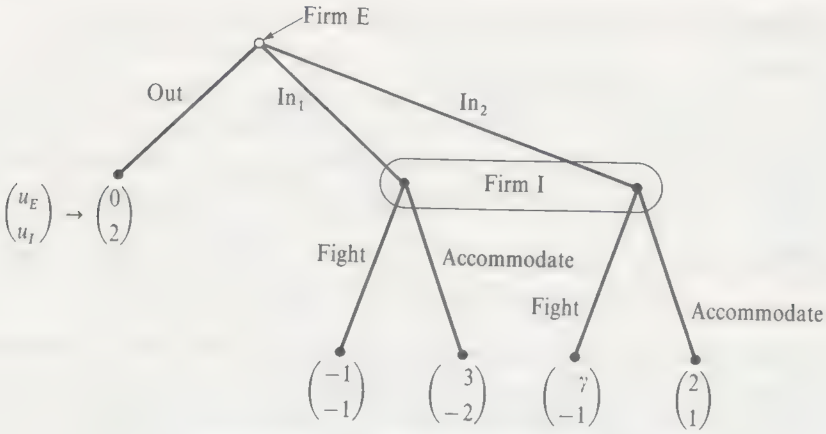
Figure 9.C.2
Extensive form for Example 9.C.2.

... must propose the joint venture since if firm E2 will accept its proposal, then ... does better proposing the joint venture than it does by either staying out or ... on its own, regardless of firm I's post-entry strategy. Next, these two ... imply that firm I's information set is reached with positive probability ... with certainty) in any weak PBE. Applying Bayesian updating at this ... on set, we conclude that the beliefs at this information set must assign a ... of 1 to being at the middle node. Given this, in any weak PBE firm I's ... must be "accommodate if entry occurs." Finally, if firm I is playing ... date if entry occurs," then firm E1 must enter if it proposes a joint venture ... firm E2 then rejects.

... nclude that the unique weak PBE in this game is a strategy–beliefs pair ... tegies of $(\sigma_{E1}, \sigma_{E2}, \sigma_I)$ = ((propose joint venture, in if E2 declines), (accept), ... odate if entry occurs)) and a belief system of μ (middle node of incumbent's ... on set) = 1. Note that this is not the only Nash equilibrium or, for that ... the only SPNE. For example, $(\sigma_{E1}, \sigma_{E2}, \sigma_I)$ = ((out, out if E2 declines), ...), (fight if entry occurs)) is an SPNE in this game. ■

... 9.C.3: In the games of Examples 9.C.1 and 9.C.2 the trick to identifying the ... PBEs consisted of seeing that some player had an optimal strategy that was ... dent of her beliefs and/or the future play of her opponents. In the game ... in Figure 9.C.3, however, this is not so for either player. Firm I is now ... fight if she thinks that firm E has played "in_1," and the optimal strategy ... firm E depends on firm I's behavior (note that $\gamma > -1$).

... ve this game, we look for a *fixed point* at which the behavior generated ... is consistent with these beliefs. We restrict attention to the case where ... Exercise 9.C.2 asks you to determine the set of weak PBEs when $\gamma \in (-1, 0)$.] ... the probability that firm I fights after entry, let μ_1 be firm I's belief that



$\gamma > -1$

"in$_1$" was E's entry strategy if entry has occurred, and let $\sigma_0, \sigma_1, \sigma_2$ denote the probabilities with which firm E actually chooses "out," "in$_1$," and "in$_2$," respectively.

Note, first, that firm I is willing to play "fight" with positive probability if and only if $-1 \geq -2\mu_1 + 1(1 - \mu_1)$, or $\mu_1 \geq \frac{2}{3}$.

Suppose, first, that $\mu_1 > \frac{2}{3}$ in a weak PBE. Then firm I must be playing "fight" with probability 1. But then firm E must be playing "in$_2$" with probability 1 (since $\gamma > 0$), and the weak PBE concept would then require that $\mu_1 = 0$, which is a contradiction.

Suppose, instead, that $\mu_1 < \frac{2}{3}$ in a weak PBE. Then firm I must be playing "accommodate" with probability 1. But, if so, then firm E must be playing "in$_1$" with probability 1, and the weak PBE concept then requires that $\mu_1 = 1$, another contradiction.

Hence, in any weak PBE of this game, we must have $\mu_1 = \frac{2}{3}$. If so, then firm E must be randomizing in the equilibrium with positive probabilities attached to both "in$_1$" and "in$_2$" and with "in$_1$" twice as likely as "in$_2$." This means that firm I's probability of playing "fight" must make firm E indifferent between "in$_1$" and "in$_2$." Hence, we must have $-1\sigma_F + 3(1 - \sigma_F) = \gamma\sigma_F + 2(1 - \sigma_F)$, or $\sigma_F = 1/(\gamma + 2)$. Firm E's payoff from playing "in$_1$" or "in$_2$" is then $(3\gamma + 2)/(\gamma + 2) > 0$, and so firm E must play "out" with zero probability. Therefore, the unique weak PBE in this game when $\gamma > 0$ has $(\sigma_0, \sigma_1, \sigma_2) = (0, \frac{2}{3}, \frac{1}{3})$, $\sigma_F = 1/(\gamma + 2)$, and $\mu_1 = \frac{2}{3}$. ■

Strengthenings of the Weak Perfect Bayesian Equilibrium Concept

We have referred to the concept defined in Definition 9.C.3 as a *weak* perfect Bayesian equilibrium because the consistency requirements that it puts on beliefs are very minimal: The *only* requirement for beliefs, other than that they specify nonnegative probabilities which add to 1 within each information set, is that they are consistent with the equilibrium strategies on the equilibrium path, in the sense of being derived from them through Bayes' rule. *No restrictions at all are placed on beliefs off the equilibrium path* (i.e., at information sets not reached with positive probability with play of the equilibrium strategies). In the literature, a number of strengthenings of this concept that put additional consistency restrictions on off-the-equilibrium-path

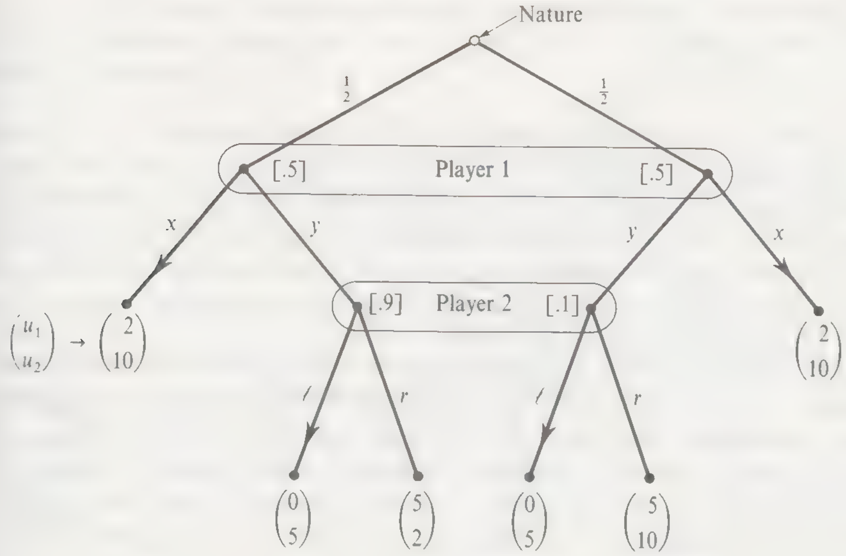


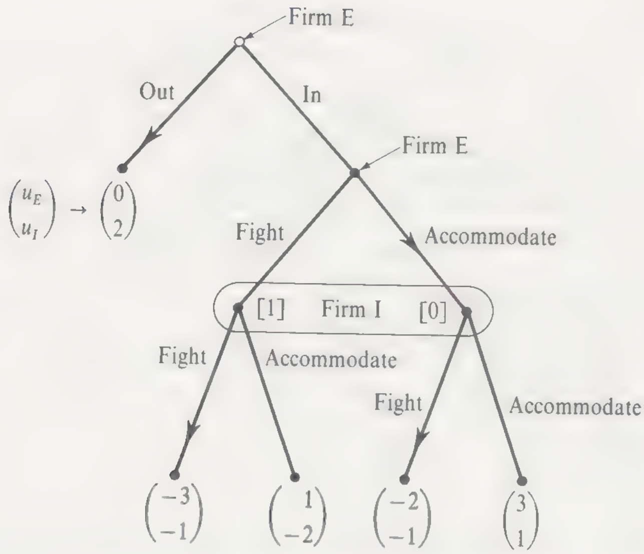
Figure 9.C.4
Extensive form for Example 9.C.4. Beliefs in a weak PBE may not be structurally consistent.

beliefs are used. Examples 9.C.4 and 9.C.5 illustrate why a strengthening of the weak PBE concept is often needed.

Example 9.C.4: Consider the game shown in Figure 9.C.4. The pure strategies and beliefs depicted in the figure constitute a weak PBE (the strategies are indicated by arrows on the chosen branches at each information set, and beliefs are indicated by numbers in brackets at the nodes in the information sets). The beliefs satisfy criterion (ii) of Definition 9.C.3; only player 1's information set is reached with positive probability, and player 1's beliefs there do reflect the probabilities assigned by nature. But the beliefs specified for player 2 in this equilibrium are not very sensible; player 2's information set can be reached only if player 1 deviates by instead choosing action y with positive probability, a deviation that must be independent of nature's actual move, since player 1 is ignorant of it. Hence, player 2 could reasonably have only beliefs that assign an equal probability to the two nodes in her information set. Here we see that it is desirable to require that beliefs at least be "structurally consistent" off the equilibrium path in the sense that there is *some* subjective probability distribution over strategy profiles that could generate probabilities consistent with the beliefs. ■

Example 9.C.5: A second and more significant problem is that a weak perfect Bayesian equilibrium need not be subgame perfect. To see this, consider again the entry game in Example 9.B.3. One weak PBE of this game involves strategies of (σ_E, σ_I) = ((out, accommodate if in), (fight if firm E plays "in")) combined with beliefs for firm I that assign probability 1 to firm E having played "fight." This weak PBE is shown in Figure 9.C.5. But note that these strategies are not subgame perfect; they do not specify a Nash equilibrium in the post-entry subgame.

The problem is that firm I's post-entry belief about firm E's post-entry play is unrestricted by the weak PBE concept because firm I's information set is off the equilibrium path. ■



Figure

These two examples indicate that the weak PBE concept can be too weak. Thus, in applications in the literature, extra consistency restrictions on beliefs are often added to the weak PBE concept to avoid these problems, with the resulting solution concept referred to as a *perfect Bayesian equilibrium.* (As a simple example, restricting attention to equilibria that induce a weak PBE in every subgame insures subgame perfection.) We shall also do this when necessary later in the book; see, in particular, the discussion of signaling in Section 13.C. For formal definitions and discussion of some notions of perfect Bayesian equilibrium, see Fudenberg and Tirole (1991a) and (1991b).

An important closely related equilibrium notion that also strengthens the weak PBE concept by embodying additional consistency restrictions on beliefs is the *sequential equilibrium* concept developed by Kreps and Wilson (1982). In contrast to notions of perfect Bayesian equilibrium (such as the one we develop in Section 13.C), the sequential equilibrium concept introduces these consistency restrictions indirectly through the formalism of a limiting sequence of strategies. Definition 9.C.4 describes its requirements.

Definition 9.C.4: A strategy profile and system of beliefs (σ, μ) is a *sequential equilibrium* of extensive form game Γ_E if it has the following properties:

(i) Strategy profile σ is sequentially rational given belief system μ.

(ii) There exists a sequence of completely mixed strategies $\{\sigma^k\}_{k=1}^{\infty}$, with $\lim_{k\to\infty} \sigma^k = \sigma$, such that $\mu = \lim_{k\to\infty} \mu^k$, where μ^k denotes the beliefs derived from strategy profile σ^k using Bayes' rule.

In essence, the sequential equilibrium notion requires that beliefs be justifiable as coming from some set of totally mixed strategies that are "close to" the equilibrium strategies σ (i.e., a small perturbation of the equilibrium strategies). This can be viewed as requiring that players can (approximately) justify their beliefs by some story in which, with some small probability, players make mistakes in choosing their strategies. Note that every sequential equilibrium is a weak perfect Bayesian equilibrium because the limiting beliefs in Definition 9.C.4 exactly coincide with the beliefs derived from the equilibrium strategies σ via Bayes' rule on the outcome path of strategy profile σ. But, in general, the reverse is not true.

As we now show, the sequential equilibrium concept strengthens the weak perfect Bayesian equilibrium concept in a manner that avoids the problems identified in Examples 9.C.4 and 9.C.5.

Example 9.C.4 Continued: Consider again the game in Figure 9.C.4. In this game, all beliefs that can be derived from any sequence of totally mixed strategies assign equal probability to the two nodes in player 2's information set. Given this fact, in any sequential equilibrium player 2 must play r and player 1 must therefore play y. The strategies (y, r) and beliefs giving equal probability to the two nodes in both players' information sets constitute the unique sequential equilibrium of this game. ■

Example 9.C.5 Continued: The unique sequential equilibrium strategies in the game of Example 9.C.5 (see Figure 9.C.5) are those of the unique SPNE: ((in, accommodate if firm E plays "in"), accommodate if firm E plays "in")). To verify this point, consider any totally mixed strategy $\bar{\sigma}$ and any node x in firm I's information set, which we denote by H_I. Letting z denote firm E's decision node following entry (the initial node of the subgame following entry), the beliefs $\mu_{\bar{\sigma}}$ associated with $\bar{\sigma}$ at information set H_I are equal to

$$\mu_{\bar{\sigma}}(x) = \frac{\text{Prob}\,(x \mid \bar{\sigma})}{\text{Prob}\,(H_I \mid \bar{\sigma})} = \frac{\text{Prob}\,(x \mid z, \bar{\sigma})\,\text{Prob}\,(z \mid \bar{\sigma})}{\text{Prob}\,(H_I \mid z, \bar{\sigma})\,\text{Prob}\,(z \mid \bar{\sigma})},$$

where $\text{Prob}\,(x \mid z, \bar{\sigma})$ is the probability of reaching node x under strategies $\bar{\sigma}$ conditional on having reached node z. Canceling terms and noting that $\text{Prob}\,(H_I \mid z, \bar{\sigma}) = 1$, we then have $\mu_{\bar{\sigma}}(x) = \text{Prob}\,(x \mid z, \bar{\sigma})$. But this is exactly the probability that firm E plays the action that leads to node x in strategy $\bar{\sigma}$. Thus, any sequence of totally mixed strategies $\{\bar{\sigma}^k\}_{k=1}^{\infty}$ that converge to σ must generate limiting beliefs for firm I that coincide with the play at node z specified in firm E's actual strategy σ_E. It is then immediate that the strategies in any sequential equilibrium must specify Nash equilibrium behavior in this post-entry subgame and thus must constitute a subgame perfect Nash equilbrium. ■

Proposition 9.C.2 gives a general result on the relation between sequential equilibria and subgame perfect Nash equilibria.

Proposition 9.C.2: In every sequential equilibrium (σ, μ) of an extensive form game Γ_E, the equilibrium strategy profile σ constitutes a subgame perfect Nash equilibrium of Γ_E.

Thus, the sequential equilibrium concept strengthens both the SPNE and the weak PBE concepts; every sequential equilibrium is both a weak PBE and an SPNE.

Although the concept of sequential equilibrium restricts beliefs that are off the equilibrium path enough to take care of the problems with the weak PBE concept illustrated in Examples 9.C.4 and 9.C.5, there are some ways in which the requirements on off-equilibrium-path beliefs embodied in the notion of sequential equilibrium may be too strong. For example, they imply that two players with the same information must have exactly the same beliefs regarding deviations by other players that have caused play to reach a given part of the game tree.

In Appendix B, we briefly describe another related (and still stronger) solution

concept, an *extensive form trembling-hand perfect Nash equilibrium*, first proposed by Selten (1975).[14]

9.D Reasonable Beliefs and Forward Induction

In Section 9.C, we saw the importance of beliefs at unreached information sets for testing the sequential rationality of a strategy. Although the weak perfect Bayesian equilibrium concept and the related stronger concepts discussed in Section 9.C can help rule out noncredible threats, in many games we can nonetheless justify a large range of off-equilibrium-path behavior by picking off-equilibrium-path beliefs appropriately (we shall see some examples shortly). This has led to a considerable amount of recent research aimed at specifying additional restrictions that "reasonable" beliefs should satisfy. In this section, we provide a brief introduction to these ideas. (We shall encounter them again when we study signaling models in Chapter 13, particularly in Appendix A of that chapter.)

To start, consider the two games depicted in Figure 9.D.1. The first is a variant of the entry game of Figure 9.C.1 in which firm I would now find it worthwhile to fight if it knew that the entrant chose strategy "in_1"; the second is a variant of the Niche Choice game of Example 9.B.4, in which firm E now targets a niche at the time of its entry. Also shown in each diagram is a weak perfect Bayesian equilibrium (arrows denote pure strategy choices, and the numbers in brackets in firm I's information set denote beliefs).

One can argue that in neither game is the equilibrium depicted very sensible.[15] Consider the game in Figure 9.D.1(a). In the weak PBE depicted, if entry occurs, firm I plays "fight" because it believes that firm E has chosen "in_1." But "in_1" is strictly dominated for firm E by "in_2." Hence, it seems reasonable to think that if firm E decided to enter, it must have used strategy "in_2." Indeed, as is commonly done in this literature, one can imagine firm E making the following speech upon entering: "I have entered, but notice that I would never have used 'in_1' to do so because 'in_2' is always a better entry strategy for me. Think about this carefully before you choose your strategy."

A similar argument holds for the weak PBE depicted in Figure 9.D.1(b). Here "small niche" is strictly dominated for firm E, not by "large niche", but by "out." Once again, firm I could not reasonably hold the beliefs that are depicted. In this case, firm I should recognize that if firm E entered rather than playing "out," it must have chosen the large niche. Now you can imagine firm E saying: "Notice that the only way I could ever do better by entering than by choosing 'out' is by targeting the large niche."

14. Selten actually gave it the name *trembling-hand perfect Nash equilibrium*; we add the modifier *extensive form* to help distinguish it from the normal form concept introduced in Section 8.F.

15. For simplicity, we focus on weak perfect Bayesian equilibria here. The points to be made apply as well to the stronger related notions discussed in Section 9.C. In fact, all the weak perfect Bayesian equilibria discussed here are also sequential equilibria; indeed, they are even extensive form trembling-hand perfect.

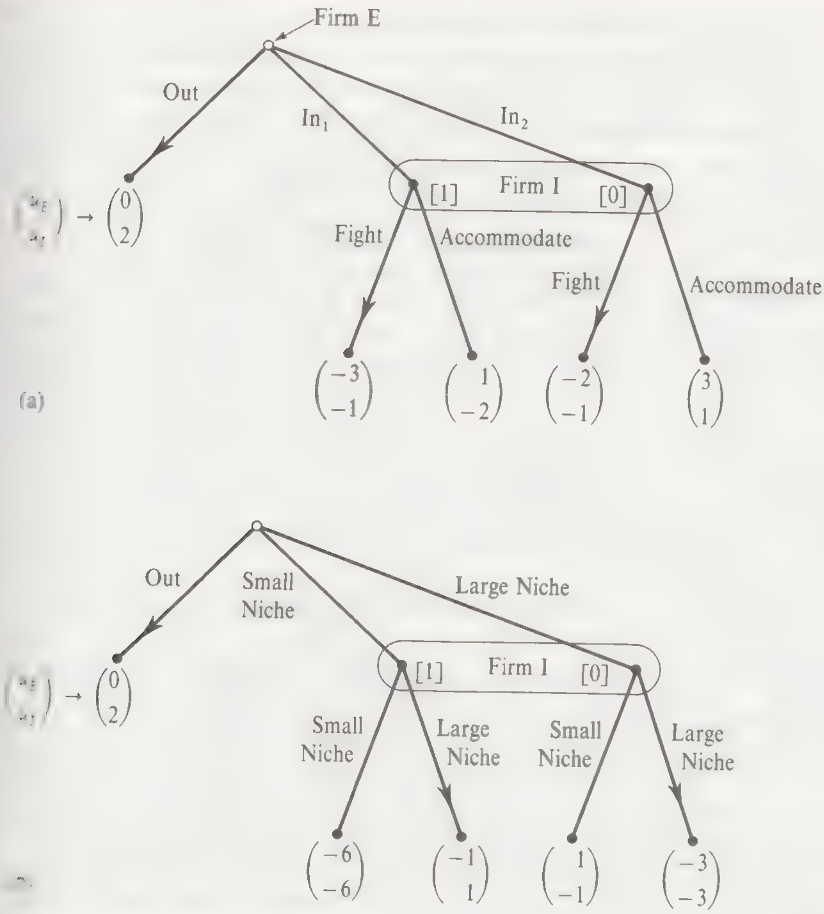


Figure 9.D.1
Two weak PBEs with unreasonable beliefs.

These arguments make use of what is known as *forward induction* reasoning [see Kohlberg (1989) and Kohlberg and Mertens (1986)]. In using backward induction, a player decides what is an optimal action for her at some point in the game tree based on her calculations of the actions that her opponents will rationally play at *later* points of the game. In contrast, in using forward induction, a player reasons about what could have rationally happened *previously*. For example, here firm I decides on its optimal post-entry action by assuming that firm E must have behaved rationally in its entry decision.

This type of idea is sometimes extended to include arguments based on *equilibrium dominance*. For example, suppose that we augment the game in Figure 9.D.1(b) by also giving firm I a move after firm E plays "out," as depicted in Figure 9.D.2 (perhaps "out" really involves entry into some alternative market of firm I's in which firm E has only one potential entry strategy).

The figure depicts a weak PBE of this game in which firm E plays "out" and firm I believes that firm E has chosen "small niche" whenever its post-entry information set is reached. In this game, "small niche" is no longer strictly dominated for firm E by "out," so our previous argument does not apply. Nevertheless, if firm E deviates from this equilibrium by entering, we can imagine firm I thinking that since firm E could have received a payoff of 0 by following its equilibrium strategy, it must be hoping to do better than that by entering, and so it must

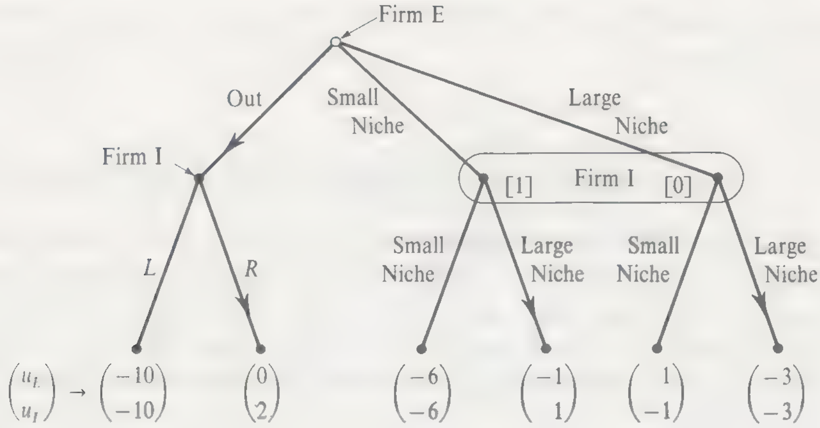


Figure 9.D.2
Strateg[...]
is equ[...]
dominate[...]

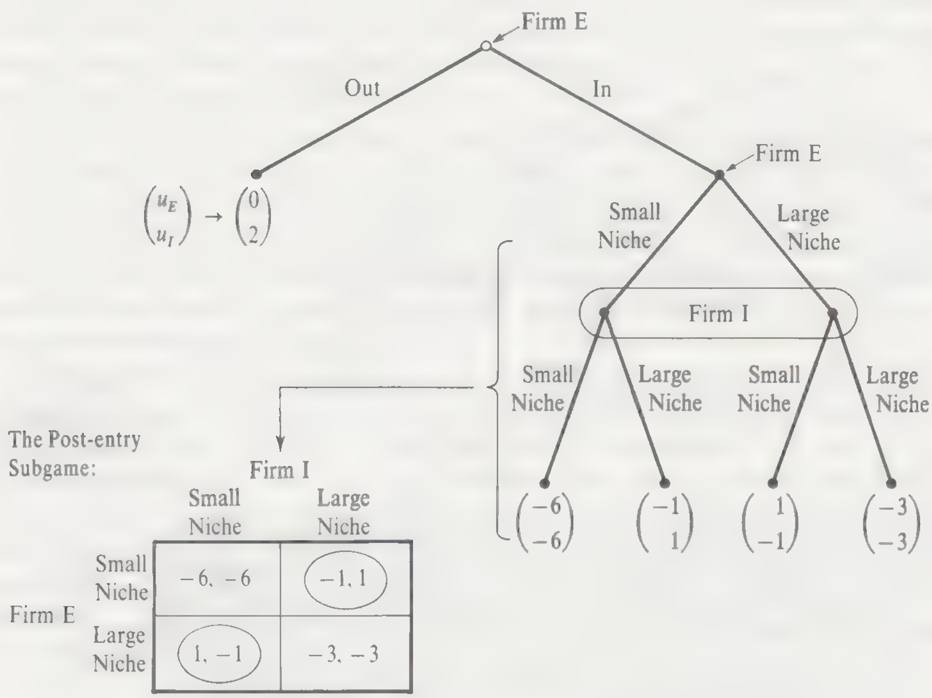


Figure 9.D.3
Forward i[...]
selects eq[...]
(large niche[...]
niche) in the [...]
subgame.

have chosen to target the large niche. In this case, we say that "small niche" is *equilibrium dominated* for firm E; that is, it is dominated if firm E treats its equilibrium payoff as something that it can achieve with certainty by following its equilibrium strategy. (This type of argument is embodied in the *intuitive criterion* refinement that we discuss in Section 13.C and Appendix A of Chapter 13 in the context of signaling models.)

Forward induction can be quite powerful. For example, reconsider the original Niche Choice game depicted in Figure 9.D.3. Recall that there are two (pure strategy) Nash equilibria in the post-entry subgame: (large niche, small niche) and (small niche, large niche). However, the force of the forward induction argument for the game in Figure 9.D.1(b) seems to apply equally well here: Strategy (in, small niche if in) is strictly dominated for firm E by playing "out." As a result, the incumbent should reason that if firm E has played "in," it intends to target the large niche in the

game. If so, firm I is better off targeting the small niche. Thus, forward ... out one of the two Nash equilibria in the post-entry subgame.

... these arguments may seem very appealing, there are also some potential ... example, suppose that we are in a world where players make mistakes ... small probability. In such a world, are the forward induction arguments ... convincing? Perhaps not. To see why, suppose that firm E enters in the ... in Figure 9.D.1(a) when it was supposed to play "out." Now firm I can ... deviation to itself as being the result of a mistake on firm E's part, a ... might equally well have led firm E to pick "in_1" as "in_2." And firm E's ... not fall on very sympathetic ears: "Of course, firm E is telling me this," ... incumbent, "it has made a mistake and now is trying to make the best ... convincing me to accommodate."

... in an even more striking manner, consider the game in Figure 9.D.3. ... firm E has entered and the two firms are about to play the simultaneous-... entry game, firm E makes its speech. But the incumbent retorts: "Forget ... just made a mistake—and even if you did not, I'm going to target ... niche!"

... the issues here, although interesting and important, are also tricky.

... feature of these forward induction arguments is how they use the normal form ... to restrict predicted play in dynamic games. This stands in sharp contrast ... earlier in this chapter, which relied exclusively on the extensive form to ... players should play in dynamic games. This raises a natural question: Can we use the normal form representation to predict play in dynamic games?

... at least two reasons why we might think we can. First, as we discussed in Chapter 7, ... as a matter of logic to think that players simultaneously choosing their ... normal form (e.g., submitting contingent plans to a referee) is equivalent to ... playing out the game dynamically as represented in the extensive form. Second, ... circumstances, it seems that the notion of weak dominance can get at the idea of ... rationality. For example, for finite games of perfect information in which no player ... at any two terminal nodes, any strategy profile surviving a process of iterated ... weakly dominated strategies leads to the same predicted outcome as the SPNE ... take a look at Example 9.B.1, and see Exercise 9.D.1).

... for using the normal form is also bolstered by the fact that extensive form ... as weak PBE can be sensitive to what may seem like irrelevant changes in the ... For example, by breaking up firm E's decision in the game in Figure 9.D.1(a) ... "in" decision followed by an "in_1" or "in_2" decision [just as we did in Figure ... game in Figure 9.D.1(b)], the unique SPNE (and, hence, the unique sequential ... becomes firm E entering and playing "in_2" and firm I accommodating. However, ... normal form associated with these two games (i.e., the normal form where we ... one of a player's strategies that have identical payoffs) is invariant to this ... extensive form; therefore, any solution based on the (reduced) normal form ... be unaffected by this change.

... points have led to a renewed interest in the use of the normal form as a device for ... in dynamic games [see, in particular, Kohlberg and Mertens (1986)]. At the ... issue remains controversial. Many game theorists believe that there is a loss ... information of strategic importance in going from the extensive form to the more ... normal form. For example, are the games in Figures 9.D.3 and 9.D.1(b) really the ... were firm I, would you be as likely to rely on the forward induction argument

in the game in Figure 9.D.3 as in that in Figure 9.D.1(b)? Does it matter for your answer whether in the game in Figure 9.D.3 a minute or a month passes between firm E's two decisions? These issues remain to be sorted out.

APPENDIX A: FINITE AND INFINITE HORIZON BILATERAL BARGAINING

In this appendix we study two models of bilateral bargaining as an economically important example of the use of the subgame perfect Nash equilibrium concept. We begin by studying a finite horizon model of bargaining and then consider its infinite horizon counterpart.

Example 9.AA.1: *Finite Horizon Bilateral Bargaining.* Two players, 1 and 2, bargain to determine the split of v dollars. The rules are as follows: The game begins in period 1; in period 1, player 1 makes an offer of a split (a real number between 0 and v) to player 2, which player 2 may then accept or reject. If she accepts, the proposed split is immediately implemented and the game ends. If she rejects, nothing happens until period 2. In period 2, the players' roles are reversed, with player 2 making an offer to player 1 and player 1 then being able to accept or reject it. Each player has a discount factor of $\delta \in (0, 1)$, so that a dollar received in period t is worth δ^{t-1} in period 1 dollars. However, after some finite number of periods T, if an agreement has not yet been reached, the bargaining is terminated and the players each receive nothing. A portion of the extensive form of this game is depicted in Figure 9.AA.1 [this model is due to Stahl (1972)].

There is a unique subgame perfect Nash equilibrium (SPNE) in this game. To see this, suppose first that T is odd, so that player 1 makes the offer in period T if no previous agreement has been reached. Now, player 2 is willing to accept *any* offer in this period because she will get zero if she refuses and the game is terminated (she is indifferent about accepting an offer of zero). Given this fact, the unique SPNE in the subgame that begins in the final period when no agreement has been previously reached has player 1 offer player 2 zero and player 2 accept.[16] Therefore, the payoffs from equilibrium play in this subgame are $(\delta^{T-1}v, 0)$.

Now consider play in the subgame starting in period $T-1$ when no previous agreement has been reached. Player 2 makes the offer in this period. In any SPNE, player 1 will accept an offer in period $T-1$ if and only if it provides her with a payoff of at least $\delta^{T-1}v$, since otherwise she will do better rejecting it and waiting to make an offer in period T (she earns $\delta^{T-1}v$ by doing so). Given this fact, in any SPNE, player 2 must make an offer in period $T-1$ that gives player 1 a payoff of exactly $\delta^{T-1}v$, and player 1 accepts this offer (note that this is player 2's best offer

16. Note that if player 2 is unwilling to accept an offer of zero, then player 1 has no optimal strategy; she wants to make a strictly positive offer ever closer to zero (since player 1 will accept any strictly positive offer). If the reliance on player 1 accepting an offer over which she is indifferent bothers you, you can convince yourself that the analysis of the game in which offers must be in small increments (pennies) yields exactly the same outcome as that identified in the text as the size of these increments goes to zero.

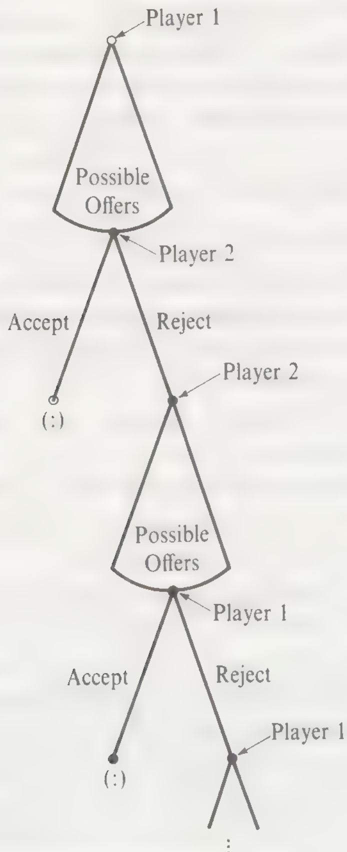


Figure 9.AA.1
The alternating-offer bilateral bargaining game.

among all those that would be accepted, and making an offer that will be rejected is worse for player 2 because it results in her receiving a payoff of zero). The payoffs arising if the game reaches period $T-1$ must therefore be $(\delta^{T-1}v, \delta^{T-2}v - \delta^{T-1}v)$.

Continuing in this fashion, we can determine that the unique SPNE when T is odd results in an agreement being reached in period 1, a payoff for player 1 of

$$\begin{aligned} v_1^*(T) &= v[1 - \delta + \delta^2 - \cdots + \delta^{T-1}] \\ &= v\left[(1-\delta)\left(\frac{1-\delta^{T-1}}{1-\delta^2}\right) + \delta^{T-1}\right], \end{aligned}$$

and a payoff to player 2 of $v_2^*(T) = v - v_1^*(T)$.

If T is instead even, then player 1 must earn $v - \delta v_1^*(T-1)$ because in any SPNE, player 2 (who will be the first offerer in the odd-number-of-periods subgame that begins in period 2 if she rejects player 1's period 1 offer) will accept an offer in period 1 if and only if it gives her at least $\delta v_1^*(T-1)$, and player 1 will offer her exactly this amount.

Finally, note that as the number of periods grows large ($T \to \infty$), player 1's payoff converges to $v/(1+\delta)$, and player 2's payoff converges to $\delta v/(1+\delta)$. ■

In Example 9.AA.1, the application of the SPNE concept was relatively straightforward; we simply needed to start at the end of the game and work backward. We now consider the infinite horizon counterpart of this game. As we noted in Section

9.B, we can no longer solve for the SPNE in this simple manner when the game has an infinite horizon. Moreover, in many games, introduction of an infinite horizon allows a broad range of behavior to emerge as subgame perfect. Nevertheless, in the infinite horizon bargaining model, the SPNE concept is quite powerful. There is a unique SPNE in this game, and it turns out to be exactly the limiting outcome of the finite horizon model as the length of the horizon T approaches ∞.

Example 9.AA.2: *Infinite Horizon Bilateral Bargaining.* Consider an extension of the finite horizon bargaining game considered in Example 9.AA.1 in which bargaining is no longer terminated after T rounds but, rather, can potentially go on forever. If this happens, the players both earn zero. This model is due to Rubinstein (1982).

We claim that this game has a unique SPNE. In this equilibrium, the players reach an immediate agreement in period 1, with player 1 earning $v/(1 + \delta)$ and player 2 earning $\delta v/(1 + \delta)$.

The method of analysis we use here, following Shaked and Sutton (1984), makes heavy use of the stationarity of the game (the subgame starting in period 2 looks exactly like that in period 1, but with the players' roles reversed).

To start, let $\bar{v}_1$ denote the largest payoff that player 1 gets in *any* SPNE (i.e., there may, in principle, be multiple SPNEs in this model).[17] Given the stationarity of the model, this is also the largest amount that player 2 can expect in the subgame that begins in period 2 after her rejection of player 1's period 1 offer, a subgame in which player 2 has the role of being the first player to make an offer. As a result, player 1's payoff in any SPNE cannot be lower than the amount $\underline{v}_1 = v - \delta\bar{v}_1$ because, if it was, then player 1 could do better by making a period 1 offer that gives player 2 just slightly more than $\delta\bar{v}_1$. Player 2 is certain to accept any such offer because she will earn only $\delta\bar{v}_1$ by rejecting it (note that we are using subgame perfection here, because we are requiring that the continuation of play after rejection is an SPNE in the continuation subgame and that player 2's response will be optimal given this fact).

Next, we claim that, in any SPNE, $\bar{v}_1$ cannot be larger than $v - \delta\underline{v}_1$. To see this, note that in any SPNE, player 2 is certain to reject any offer in period 1 that gives her less than $\delta\underline{v}_1$ because she can earn at least $\delta\underline{v}_1$ by rejecting it and waiting to make an offer in period 2. Thus, player 1 can do no better than $v - \delta\underline{v}_1$ by making an offer that is accepted in period 1. What about by making an offer that is rejected in period 1? Since player 2 must earn at least $\delta\underline{v}_1$ if this happens, and since agreement cannot occur before period 2, player 1 can earn no more than $\delta v - \delta\underline{v}_1$ by doing this. Hence, we have $\bar{v}_1 \leq v - \delta\underline{v}_1$.

Next, note that these derivations imply that

$$\bar{v}_1 \leq v - \delta\underline{v}_1 = (\underline{v}_1 + \delta\bar{v}_1) - \delta\underline{v}_1,$$

so that

$$\bar{v}_1(1 - \delta) \leq \underline{v}_1(1 - \delta).$$

Given the definitions of $\underline{v}_1$ and $\bar{v}_1$, this implies that $\underline{v}_1 = \bar{v}_1$, and so player 1's SPNE payoff is uniquely determined. Denote this payoff by v_1°. Since $v_1^\circ = v - \delta v_1^\circ$, we find that player 1 must earn $v_1^\circ = v/(1 + \delta)$ and player 2 must earn $v_2^\circ = v - v_1^\circ = \delta v/(1 + \delta)$. In addition, recalling the argument in the previous paragraph, we see

17. This maximum can be shown to be well defined, but we will not do so here.

that an agreement will be reached in the first period (player 1 will find it worthwhile to make an offer that player 2 accepts). The SPNE strategies are as follows: A player who has just received an offer accepts it if and only if she is offered at least δv_1°, while a player whose turn it is to make an offer offers exactly δv_1° to the player receiving the offer.

Note that the equilibrium strategies, outcome, and payoffs are precisely the limit of those in the finite game in Example 9.AA.1 as $T \to \infty$. ■

The coincidence of the infinite horizon equilibrium with the limit of the finite horizon equilibria in this model is not a general property of infinite horizon games. The discussion of infinitely repeated games in Chapter 12 provides an illustration of this point.

We should also point out that the outcomes of game-theoretic models of bargaining can be quite sensitive to the precise specification of the bargaining process and players' preferences. Exercises 9.B.7 and 9.B.13 provide an illustration.

APPENDIX B: EXTENSIVE FORM TREMBLING-HAND PERFECT NASH EQUILIBRIUM

In this appendix we extend the analysis presented in Section 9.C by discussing another equilibrium notion that strengthens the consistency conditions on beliefs in the weak PBE concept: *extensive form trembling-hand perfect Nash equilibrium* [due to Selten (1975)]. In fact, this equilibrium concept is the strongest among those discussed in Section 9.C.

The definition of an extensive form trembling-hand perfect Nash equilibrium parallels that for the normal form (see Section 8.F) but has the trembles applied not to a player's mixed strategies, but rather to the player's choice at each of her information sets. A useful way to view this idea is with what Selten (1975) calls the *agent normal form*. This is the normal form that we would derive if we pretended that the player had a set of agents in charge of moving for her at each of her information sets (a different one for each), each acting independently to try to maximize the player's payoff.

Definition 9.BB.1: Strategy profile σ in extensive form game Γ_E is an *extensive form trembling-hand perfect Nash equilibrium* if and only if it is a normal form trembling-hand perfect Nash equilibrium of the agent normal form derived from Γ_E.

To see why it is desirable to have the trembles occurring at each information set rather than over strategies as in the normal-form concept considered in Section 8.F, consider Figure 9.BB.1, which is taken from van Damme (1983). This game has a unique subgame perfect Nash equilibrium: $(\sigma_1, \sigma_2) = ((NR, L), \ell)$. But you can check that $((NR, L), \ell)$ is not the only normal form trembling-hand perfect Nash equilibrium: so are $((R, L), r)$ and $((R, M), r)$. The reason that these two strategy profiles are normal form trembling-hand perfect is that, in the normal form, the tremble to strategy (NR, M) by player 1 can be larger than that to (NR, L) despite the fact that the latter is a better choice for player 1 at her second decision node.

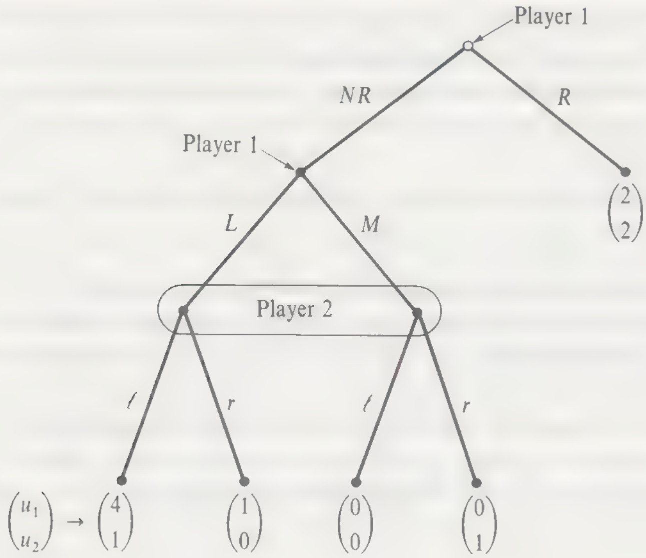


Figure
Str
((R.
((R.
form
perf
subg

With such a tremble, player 2's best response to player 1's perturbed strategy is r. It is not difficult to see, however, that the unique extensive form trembling-hand perfect Nash equilibrium of this game is $((NR, L), \ell)$ because the agent who moves at player 1's second decision node will put as high a probability as possible on L.

When we compare Definitions 9.BB.1 and 9.C.4, it is apparent that every extensive form trembling-hand perfect Nash equilibrium is a sequential equilibrium. In particular, even though the trembling-hand perfection criterion is not formulated in terms of beliefs, we can use the sequence of (strictly mixed) equilibrium strategies $\{\sigma^k\}_{k=1}^{\infty}$ in the perturbed games of the agent normal form as our strategy sequence for deriving sequential equilibrium beliefs. Because the limiting strategies σ in the extensive form trembling-hand perfect equilibrium are best responses to every element of this sequence, they are also best responses to each other with these derived beliefs. (Every extensive form trembling-hand perfect Nash equilibrium is therefore also subgame perfect.)

In essence, by introducing trembles, the extensive form trembling-hand perfect equilibrium notion makes every part of the tree be reached when strategies are perturbed, and because equilibrium strategies are required to be best responses to perturbed strategies, it insures that equilibrium strategies are sequentially rational. The primary difference between this notion and that of sequential equilibrium is that, like its normal form cousin, the extensive form trembling-hand perfect equilibrium concept can also eliminate some sequential equilibria in which weakly dominated strategies are played. Figure 9.BB.2 (a slight modification of the game in Figure 9.C.1) depicts a sequential equilibrium whose strategies are not extensive form trembling-hand perfect.

In general, however, the concepts are quite close [see Kreps and Wilson (1982) for a formal comparison]; and because it is much easier to check that strategies are best responses at the limiting beliefs than it is to check that they are best responses for a sequence of strategies, sequential equilibrium is much more commonly used. For an interesting further discussion of this concept, consult van Damme (1983).

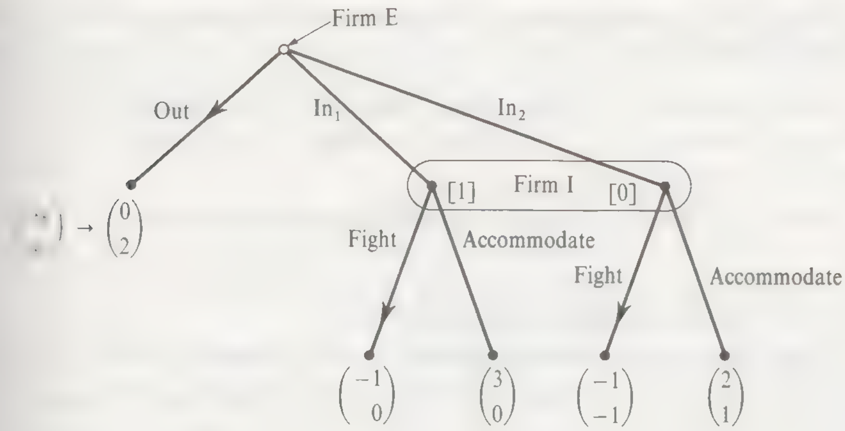


Figure 9.BB.2
A sequential equilibrium need not be extensive form trembling-hand perfect

ERENCES

B. D. (1984). Rationalizable strategic behavior. *Econometrica* **52**: 1007–28.

M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

erg, D., and J. Tirole. (1991a). Perfect Bayesian and sequential equilibrium. *Journal of Economic Theory* **53**: 236–60.

rg, D., and J. Tirole. (1991b). *Game Theory*. Cambridge, Mass.: MIT Press.

E. (1989). Refinement of Nash equilibrium: the main ideas. Harvard Business School Working Paper No. 89-073.

rg, E., and J.-F. Mertens. (1986). On the strategic stability of equilibria. *Econometrica* **54**: 1003–38.

D. M., and R. Wilson. (1982). Sequential equilibrium. *Econometrica* **50**: 863–94.

H. (1981). *Game Theory for the Social Sciences*. New York: New York University Press.

R. (1991). *Game Theory: Analysis of Conflict*. Cambridge, Mass.: Harvard University Press.

D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica* **52**: 29–50.

R. (1981). Games of perfect information, predatory pricing, and the chain-store paradox. *Journal of Economic Theory* **25**: 92–100.

A. (1982). Perfect equilibrium in a bargaining model. *Econometrica* **50**: 97–109.

R. (1965). Spieltheoretische behandlung eines oligopolmodells mit nachfragetragheit. *Zeitschrift für die gesamte Staatswissenschaft* **121**: 301–24.

R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **4**: 25–55.

A., and J. Sutton. (1984). Involuntary unemployment as a perfect equilibrium in a bargaining model. *Econometrica* **52**: 1351–64.

(1972). *Bargaining Theory*. Stockholm: Economics Research Unit.

E. (1983). *Refinements of the Nash Equilibrium Concept*. Berlin: Springer-Verlag.

ERCISES

How many subgames are there in the game of Example 9.B.2 (depicted in Figure 9.B.3)?

In text.

Verify that the strategies identified through backward induction in Example 9.B.2 te a Nash equilibrium of the game studied there. Also, identify *all other* pure strategy equilibria of this game. Argue that each of these other equilibria does not satisfy the of sequential rationality.

9.B.4$^{\mathrm{B}}$ Prove that in a finite *zero-sum* game of perfect information, there are unique subgame perfect Nash equilibrium payoffs.

9.B.5$^{\mathrm{B}}$ (E. Maskin) Consider a game with two players, player 1 and player 2, in which each player i can choose an action from a finite set M_i that contains m_i actions. Player i's payoff if the action choices are (m_1, m_2) is $\phi_i(m_1, m_2)$.

(a) Suppose, first, that the two players move simultaneously. How many strategies does each player have?

(b) Now suppose that player 1 moves first and that player 2 observes player 1's move before choosing her move. How many strategies does each player have?

(c) Suppose that the game in **(b)** has multiple SPNEs. Show that if this is the case, then there exist two pairs of moves (m_1, m_2) and (m_1', m_2') (where either $m_1 \neq m_1'$ or $m_2 \neq m_2'$) such that either

$$\text{(i)}\quad \phi_1(m_1, m_2) = \phi_1(m_1', m_2')$$

or

$$\text{(ii)}\quad \phi_2(m_1, m_2) = \phi_2(m_1', m_2').$$

(d) Suppose that for any two pairs of moves (m_1, m_2) and (m_1', m_2') such that $m_1 \neq m_1'$ or $m_2 \neq m_2'$, condition (ii) is violated (i.e., player 2 is never indifferent between pairs of moves). Suppose also that there exists a pure strategy Nash equilibrium in the game in **(a)** in which π_1 is player 1's payoff. Show that in any SPNE of the game in **(b)**, player 1's payoff is at least π_1. Would this conclusion necessarily hold for any *Nash* equilibrium of the game in **(b)**?

(e) Show by example that the conclusion in **(d)** may fail either if condition (ii) holds for some strategy pairs (m_1, m_2), (m_1', m_2') with $m_1 \neq m_1'$ or $m_2 \neq m_2'$ or if we replace the phrase *pure strategy Nash equilibrium* with the phrase *mixed strategy Nash equilibrium*.

9.B.6$^{\mathrm{B}}$ Solve for the mixed strategy equilibrium involving actual randomization in the post-entry subgame of the Niche Choice game in Example 9.B.4. Is there an SPNE that induces this behavior in the post-entry subgame? What are the SPNE strategies?

9.B.7$^{\mathrm{B}}$ Consider the finite horizon bilateral bargaining game in Appendix A (Example 9.AA.1); but instead of assuming that players discount future payoffs, assume that it costs $c < v$ to make an offer. (Only the player making an offer incurs this cost, and players who have made offers incur this cost even if no agreement is ultimately reached.) What is the (unique) SPNE of this alternative model? What happens as T approaches ∞?

9.B.8$^{\mathrm{C}}$ Prove that every (finite) game Γ_E has a mixed strategy subgame perfect Nash equilibrium.

9.B.9$^{\mathrm{B}}$ Consider a game in which the following simultaneous-move game is played twice:

		Player 2		
		b_1	b_2	b_3
	a_1	10, 10	2, 12	0, 13
Player 1	a_2	12, 2	5, 5	0, 0
	a_3	13, 0	0, 0	1, 1

The players observe the actions chosen in the first play of the game prior to the second play. What are the pure strategy subgame perfect Nash equilibria of this game?

9.B.10$^{\text{B}}$ Reconsider the game in Example 9.B.3, but now change the post-entry game so that when both players choose "accommodate", instead of receiving the payoffs $(u_E, u_I) = (3, 1)$, the players now must play the following simultaneous-move game:

		Firm I	
		l	r
Firm E	U	3, 1	0, 0
	D	0, 0	x, 3

What are the SPNEs of this game when $x \geq 0$? When $x < 0$?

9.B.11$^{\text{B}}$ Two firms, A and B, are in a market that is declining in size. The game starts in period 0 and the firms can compete in periods 0, 1, 2, 3, ... (i.e., indefinitely) if they so choose. Duopoly profits in period t for firm A are equal to $105 - 10t$, and they are $10.5 - t$ for firm B. Monopoly profits (those if a firm is the only one left in the market) are $510 - 25t$ for firm A and $51 - 2t$ for firm B.

Suppose that at the start of each period, each firm must decide either to "stay in" or "exit" if it is still active (they do so simultaneously if both are still active). Once a firm exits, it is out of the market forever and earns zero in each period thereafter. Firms maximize their (undiscounted) sum of profits.

What is this game's subgame perfect Nash equilibrium outcome (and what are the firms' strategies in the equilibrium)?

9.B.12$^{\text{C}}$ Consider the infinite horizon bilateral bargaining model of Appendix A (Example 9.AA.2). Suppose the discount factors δ_1 and δ_2 of the two players differ. Now what is the (unique) subgame perfect Nash equilibrium?

9.B.13$^{\text{B}}$ What are the subgame perfect Nash equilibria of the infinite horizon version of Exercise 9.B.7?

9.B.14$^{\text{B}}$ At time 0, an incumbent firm (firm I) is already in the widget market, and a potential entrant (firm E) is considering entry. In order to enter, firm E must incur a cost of $K > 0$. Firm E's only opportunity to enter is at time 0. There are three production periods. In any period in which both firms are active in the market, the game in Figure 9.Ex.1 is played. Firm E moves first, deciding whether to stay in or exit the market. If it stays in, firm I decides whether to fight (the upper payoff is for firm E). Once firm E plays "out," it is out of

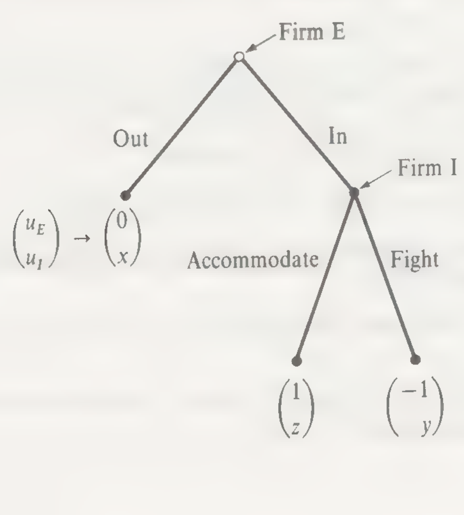


Figure 9.Ex.1

the market forever; firm E earns zero in any period during which it is out of the market, and firm I earns x. The discount factor for both firms is δ.

Assume that:

(A.1) $x > z > y$.
(A.2) $y + \delta x > (1 + \delta)z$.
(A.3) $1 + \delta > K$.

(a) What is the (unique) subgame perfect Nash equilibrium of this game?

(b) Suppose now that firm E faces a financial constraint. In particular, if firm I fights *once* against firm E (in any period), firm E will be forced out of the market from that point on. Now what is the (unique) subgame perfect Nash equilibrium of this game? (If the answer depends on the values of parameters beyond the three assumptions, indicate how.)

9.C.1^B Prove Proposition 9.C.1.

9.C.2^B What is the set of weak PBEs in the game in Example 9.C.3 when $\gamma \in (-1, 0)$?

9.C.3^C A buyer and a seller are bargaining. The seller owns an object for which the buyer has value $v > 0$ (the seller's value is zero). This value is known to the buyer but not to the seller. The value's prior distribution is common knowledge. There are two periods of bargaining. The seller makes a take-it-or-leave-it offer (i.e., names a price) at the start of each period that the buyer may accept or reject. The game ends when an offer is accepted or after two periods, whichever comes first. Both players discount period 2 payoffs with a discount factor of $\delta \in (0, 1)$.

Assume throughout that the buyer always accepts the seller's offer whenever she is indifferent.

(a) Characterize the (pure strategy) weak perfect Bayesian equilibria for a case in which v can take two values v_L and v_H, with $v_H > v_L > 0$, and where $\lambda = \text{Prob}(v_H)$.

(b) Do the same for the case in which v is uniformly distributed on $[\underline{v}, \bar{v}]$.

9.C.4^C A plaintiff, Ms. P, files a suit against Ms. D (the defendant). If Ms. P wins, she will collect π dollars in damages from Ms. D. Ms. D knows the likelihood that Ms. P will win, $\lambda \in [0, 1]$, but Ms. P does not (Ms. D might know if she was actually at fault). They both have strictly positive costs of going to trial of c_p and c_d. The prior distribution of λ has density $f(\lambda)$ (which is common knowledge).

Suppose pretrial settlement negotiations work as follows: Ms. P makes a take-it-or-leave-it settlement offer (a dollar amount) to Ms. D. If Ms. D accepts, she pays Ms. P and the game is over. If she does not accept, they go to trial.

(a) What are the (pure strategy) weak perfect Bayesian equilibria of this game?

(b) What effects do changes in c_p, c_d, and π have?

(c) Now allow Ms. D, after having her offer rejected, to decide not to go to court after all. What are the weak perfect Bayesian equilibria? What about the effects of the changes in **(b)**?

9.C.5^C Reconsider Exercise 9.C.4. Now suppose it is Ms. P who knows λ.

9.C.6^B What are the sequential equilibria in the games in Exercises 9.C.3 to 9.C.5?

9.C.7^B (Based on work by K. Bagwell and developed as an exercise by E. Maskin) Consider the extensive form game depicted in Figure 9.Ex.2.

(a) Find a subgame perfect Nash equilibrium of this game. Is it unique? Are there any other Nash equilibria?

(b) Now suppose that player 2 cannot observe player 1's move. Write down the new extensive form. What is the set of Nash equilibria?

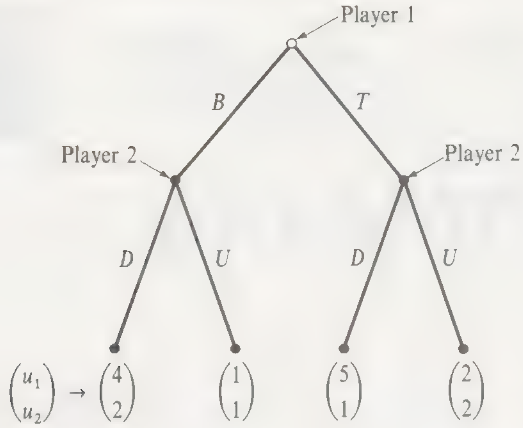


Figure 9.Ex.2

(c) Now suppose that player 2 observes player 1's move correctly with probability $p \in (0, 1)$ and incorrectly with probability $1 - p$ (e.g., if player 1 plays T, player 2 observes T with probability p and observes B with probability $1 - p$). Suppose that player 2's propensity to observe incorrectly (i.e., given by the value of p) is common knowledge to the two players. What is the extensive form now? Show that there is a unique weak perfect Bayesian equilibrium. What is it?

9.D.1^{B} Show that under the condition given in Proposition 9.B.2 for existence of a unique subgame perfect Nash equilibrium in a finite game of perfect information, there is an order of iterated removal of weakly dominated strategies for which all surviving strategy profiles lead to the same outcome (i.e., have the same equilibrium path and payoffs) as the subgame perfect Nash equilibrium. [In fact, *any* order of deletion leads to this result; see Moulin (1981).]

Market Equilibrium and Market Failure

In Part III, our focus shifts to the fundamental issue of economics: *the organization of production and the allocation of the resulting commodities among consumers*. This fundamental issue can be addressed from two perspectives, one *positive* and the other *normative*.

From a positive (or *descriptive*) perspective, we can investigate the determination of production and consumption under various institutional mechanisms. The institutional arrangement that is our central focus is that of a *market* (or *private ownership*) *economy*. In a market economy, individual consumers have ownership rights to various assets (such as their labor) and are free to trade these assets in the marketplace for other assets or goods. Likewise, firms, which are themselves owned by consumers, decide on their production plan and trade in the market to secure necessary inputs and sell the resulting outputs. Roughly speaking, we can identify a *market equilibrium* as an outcome of a market economy in which each agent in the economy (i.e., each consumer and firm) is doing as well as he can given the actions of all other agents.

In contrast, from a normative (or *prescriptive*) perspective, we can ask what constitutes a *socially optimal* plan of production and consumption (of course, we will need to be more specific about what "socially optimal" means), and we can then examine the extent to which specific institutions, such as a market economy, perform well in this regard.

In Chapter 10, we study *competitive* (or *perfectly competitive*) *market economies* for the first time. These are market economies in which every relevant good is traded in a market at publicly known prices and all agents act as price takers (recall that much of the analysis of individual behavior in Part I was geared to this case). We begin by defining, in a general way, two key concepts: *competitive* (or *Walrasian*) *equilibrium* and *Pareto optimality* (or *Pareto efficiency*). The concept of competitive equilibrium provides us with an appropriate notion of market equilibrium for competitive market economies. The concept of Pareto optimality offers a minimal and uncontroversial test that any social optimal economic outcome should pass. An economic outcome is said to be Pareto optimal if it is impossible to make some individuals better off without making some other individuals worse off. This concept is a formalization of the idea that there is no waste in society, and it conveniently

separates the issue of economic efficiency from more controversial (and p[illegible]) questions regarding the ideal *distribution* of well-being across individuals.

Chapter 10 then explores these two concepts and the relationships between them in the special context of the *partial equilibrium model*. The partial equilibrium model, which forms the basis for our analysis throughout Part III, offers a considerable analytical simplification; in it, our analysis can be conducted by analyzing a single market (or a small group of related markets) at a time. In this special context, we establish two central results regarding the optimality properties of competitive equilibria, known as the *fundamental theorems of welfare economics*. These can be roughly paraphrased as follows:

> ***The First Fundamental Welfare Theorem.*** If every relevant good is traded in a market at publicly known prices (i.e., if there is a complete set of markets), and if households and firms act perfectly competitively (i.e., as price takers), then the market outcome is Pareto optimal. That is, when markets are complete, any *competitive equilibrium is necessarily Pareto optimal.*

> ***The Second Fundamental Welfare Theorem.*** If household preferences and firm production sets are convex, there is a complete set of markets with publicly known prices, and every agent acts as a price taker, then *any Pareto optimal outcome can be achieved as a competitive equilibrium if appropriate lump-sum transfers of wealth are arranged.*

The first welfare theorem provides a set of conditions under which we can be assured that a market economy will achieve a Pareto optimal result; it is, in a sense, the formal expression of Adam Smith's claim about the "invisible hand" of the market. The second welfare theorem goes even further. It states that under the same set of assumptions as the first welfare theorem plus convexity conditions, *all* Pareto optimal outcomes can in principle be implemented through the market mechanism. That is, a public authority who wishes to implement a particular Pareto optimal outcome (reflecting, say, some political consensus on proper distributional goals) may always do so by appropriately redistributing wealth and then "letting the market work."

In an important sense, the first fundamental welfare theorem establishes the perfectly competitive case as a benchmark for thinking about outcomes in market economies. In particular, any inefficiencies that arise in a market economy, and hence any role for Pareto-improving market intervention, *must* be traceable to a violation of at least one of the assumptions of this theorem.

The remainder of Part III, Chapters 11 to 14, can be viewed as a development of this theme. In these chapters, we study a number of ways in which actual markets may depart from this perfectly competitive ideal and where, as a result, market equilibria fail to be Pareto optimal, a situation known as *market failure*.

In Chapter 11, we study *externalities* and *public goods*. In both cases, the actions of one agent directly affect the utility functions or production sets of other agents in the economy. We see there that the presence of these nonmarketed "goods" or "bads" (which violates the complete markets assumption of the first welfare theorem) undermines the Pareto optimality of market equilibrium.

In Chapter 12, we turn to the study of settings in which some agents in the economy have *market power* and, as a result, fail to act as price takers. Once again,

...ption of the first fundamental welfare theorem fails to hold, and market ... fail to be Pareto optimal as a result.

... Chapters 13 and 14, we consider situations in which an *asymmetry of* ... exists among market participants. The complete markets assumption of ... welfare theorem implicitly requires that the characteristics of traded ...es be observable by all market participants because, without this observ- ...inct markets cannot exist for commodities that have different characteristics. ... 13 focuses on the case in which asymmetric information exists between agents ... of contracting. Our discussion highlights several phenomena—*adverse* ... *signaling*, and *screening*—that can arise as a result of this informational ...n, and the welfare loss that it causes. Chapter 14 in contrast, investigates ... of postcontractual asymmetric information, a problem that leads us to the ... *principal-agent model*. Here, too, the presence of asymmetric information ...rade of all relevant commodities and can lead market outcomes to be Pareto ...t.

... extensively in some places in Part III on the tools that we developed in ... II. This is particularly true in Chapter 10, where we use material developed ... and Chapters 12 and 13, where we use the game-theoretic tools developed ... II.

... more complete and general study of competitive market economies and ...mental welfare theorems is reserved for Part IV.

Competitive Markets

Introduction

...chapter, we consider, for the first time, an entire economy in which consumers ...rms interact through markets. The chapter has two principal goals: first, to ...ly introduce and study two key concepts, the notions of *Pareto optimality* and ...*itive equilibrium*, and second, to develop a somewhat special but analytically ...ractable context for the study of market equilibrium, the *partial equilibrium* ...

...e begin in Section 10.B by presenting the notions of a *Pareto optimal* (or *Pareto* ... *allocation* and of a *competitive* (or *Walrasian*) *equilibrium* in a general ...

...arting in Section 10.C, we narrow our focus to the partial equilibrium context. ...artial equilibrium approach, which originated in Marshall (1920), envisions the ... for a single good (or group of goods) for which each consumer's expenditure ...tes only a small portion of his overall budget. When this is so, it is ...able to assume that changes in the market for this good will leave the prices ...ther commodities approximately unaffected and that there will be, in addition, ...e wealth effects in the market under study. We capture these features in the ... possible way by considering a two-good model in which the expenditure on ...modities other than that under consideration is treated as a single composite ...dity (called the *numeraire* commodity), and in which consumers' utility ... take a quasilinear form with respect to this numeraire. Our study of the ...ive equilibria of this simple model lends itself to extensive demand-and-...graphical analysis. We also discuss how to determine the comparative statics ... that arise from exogenous changes in the market environment. As an ...n, we consider the effects on market equilibrium arising from the introduc-... a distortionary commodity tax.

...ection 10.D, we analyze the properties of Pareto optimal allocations in the ...equilibrium model. Most significantly, we establish for this special context the ... the *fundamental theorems of welfare economics*: Competitive equilibrium ... are necessarily Pareto optimal, and any Pareto optimal allocation can ...ed as a competitive equilibrium if appropriate lump-sum transfers are made.

As we noted in the introduction to Part III, these results identify an important benchmark case in which market equilibria yield desirable economic outcomes. At the same time, they provide a framework for identifying situations of market failure, such as those we study in Chapters 11 to 14.

In Section 10.E, we consider the measurement of welfare changes in the partial equilibrium context. We show that these can be represented by areas between properly defined demand and supply curves. As an application, we examine the deadweight loss of distortionary taxation.

Section 10.F contemplates settings characterized by *free entry*, that is, settings in which all potential firms have access to the most efficient technology and may enter and exit markets in response to the profit opportunities they present. We define a notion of *long-run competitive equilibrium* and then use it to distinguish between long-run and short-run comparative static effects in response to changes in market conditions.

In Section 10.G, we provide a more extended discussion of the use of partial equilibrium analysis in economic modeling.

The material covered in this chapter traces its roots far back in economic thought. An excellent source for further reading is Stigler (1987). We should emphasize that the analysis of competitive equilibrium and Pareto optimality presented here is very much a first pass. In Part IV we return to the topic for a more complete and general investigation; many additional references will be given there.

10.B Pareto Optimality and Competitive Equilibria

In this section, we introduce and discuss the concepts of *Pareto optimality* (or *Pareto efficiency*) and *competitive* (or *Walrasian*) *equilibrium* in a general setting.

Consider an economy consisting of I consumers (indexed by $i = 1, \dots, I$), J firms (indexed by $j = 1, \dots, J$), and L goods (indexed by $\ell = 1, \dots, L$). Consumer i's preferences over consumption bundles $x_i = (x_{1i}, \dots, x_{Li})$ in his consumption set $X_i \subset \mathbb{R}^L$ are represented by the utility function $u_i(\cdot)$. The total amount of each good $\ell = 1, \dots, L$ initially available in the economy, called the total *endowment* of good ℓ, is denoted by $\omega_\ell \geq 0$ for $\ell = 1, \dots, L$. It is also possible, using the production technologies of the firms, to transform some of the initial endowment of a good into additional amounts of other goods. Each firm j has available to it the production possibilities summarized by the production set $Y_j \subset \mathbb{R}^L$. An element of Y_j is a production vector $y_j = (y_{1j}, \dots, y_{Lj}) \in \mathbb{R}^L$. Thus, if $(y_1, \dots, y_J) \in \mathbb{R}^{LJ}$ are the production vectors of the J firms, the total (net) amount of good ℓ available to the economy is $\omega_\ell + \sum_j y_{\ell j}$ (recall that negative entries in a production vector denote input usage; see Section 5.B).

We begin with Definition 10.B.1, which identifies the set of possible outcomes in this economy:

Definition 10.B.1: An *economic allocation* $(x_1, \dots, x_I, y_1, \dots, y_J)$ is a specification of a consumption vector $x_i \in X_i$ for each consumer $i = 1, \dots, I$ and a production vector $y_j \in Y_j$ for each firm $j = 1, \dots, J$. The allocation $(x_1, \dots, x_I, y_1, \dots, y_J)$ is *feasible* if

$$\sum_{i=1}^{I} x_{\ell i} \leq \omega_\ell + \sum_{j=1}^{J} y_{\ell j} \quad \text{for } \ell = 1, \dots, L.$$

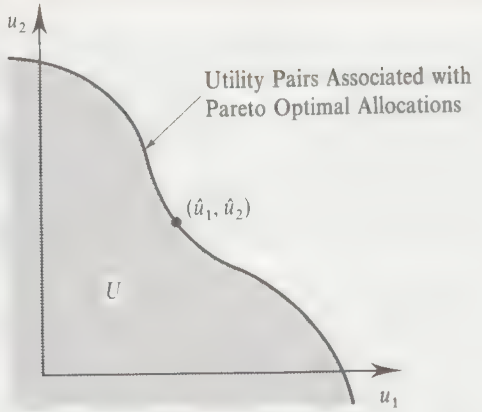


Figure 10.B.1
A utility possibility set.

... economic allocation is feasible if the total amount of each good ... does not exceed the total amount available from both the initial ... and production.

...ality

... interest to ask whether an economic system is producing an "optimal" ...come. An essential requirement for any optimal economic allocation ...ssess the property of *Pareto optimality* (or *Pareto efficiency*).

... 2 A feasible allocation $(x_1, \ldots, x_I, y_1, \ldots, y_J)$ is *Pareto optimal* (or ...*cient*) if there is no other feasible allocation $(x'_1, \ldots, x'_I, y'_1, \ldots, y'_J)$ ... $u_i(x'_i) \geq u_i(x_i)$ for all $i = 1, \ldots, I$ and $u_i(x'_i) > u_i(x_i)$ for some i.

...n that is Pareto optimal uses society's initial resources and techno-...ities efficiently in the sense that there is no alternative way to organize ... and distribution of goods that makes some consumer better off ...ing some other consumer worse off.

...B.1 illustrates the concept of Pareto optimality. There we depict the set ...tility levels in a two-consumer economy. This set is known as a *utility* ... and is defined in this two-consumer case by

$$\ldots (u_1, u_2) \in \mathbb{R}^2\text{: there exists a feasible allocation } (x_1, x_2, y_1, \ldots, y_J)$$
$$\text{such that } u_i \leq u_i(x_i) \text{ for } i = 1, 2\}.$$

...o optimal allocations corresponds to those allocations that generate ...g in the utility possibility set's northeast boundary, such as point ... such point, it is impossible to make one consumer better off without ... other worse off.

...rtant to note that the criterion of Pareto optimality does not insure ...ation is in any sense equitable. For example, using all of society's ... technological capabilities to make a single consumer as well off as ...ect to all other consumers receiving a subsistence level of utility, results ...cation that is Pareto optimal but not in one that is very desirable on ...al grounds. Nevertheless, Pareto optimality serves as an important ...r the desirability of an allocation; it does, at the very least, say that ...waste in the allocation of resources in society.

Competitive Equilibria

Throughout this chapter, we are concerned with the analysis of competitive market economies. In such an economy, society's initial endowments and technological possibilities (i.e., the firms) are owned by consumers. We suppose that consumer i initially owns $\omega_{\ell i}$ of good ℓ, where $\sum_i \omega_{\ell i} = \omega_\ell$. We denote consumer i's vector of endowments by $\omega_i = (\omega_{1i}, \ldots, \omega_{Li})$. In addition, we suppose that consumer i owns a share θ_{ij} of firm j (where $\sum_i \theta_{ij} = 1$), giving him a claim to fraction θ_{ij} of firm j's profits.

In a competitive economy, a market exists for each of the L goods, and all consumers and producers act as price takers. The idea behind the price-taking assumption is that if consumers and producers are small relative to the size of the market, they will regard market prices as unaffected by their own actions.[1]

Denote the vector of market prices for goods $1, \ldots, L$ by $p = (p_1, \ldots, p_L)$. Definition 10.B.3 introduces the notion of a competitive (or Walrasian) equilibrium.

Definition 10.B.3: The allocation $(x_1^*, \ldots, x_I^*, y_1^*, \ldots, y_J^*)$ and price vector $p^* \in \mathbb{R}^L$ constitute a *competitive* (or *Walrasian*) *equilibrium* if the following conditions are satisfied:

(i) *Profit maximization*: For each firm j, y_j^* solves

$$\underset{y_j \in Y_j}{\text{Max}} \quad p^* \cdot y_j. \tag{10.B.1}$$

(ii) *Utility maximization*: For each consumer i, x_i^* solves

$$\begin{aligned} &\underset{x_i \in X_i}{\text{Max}} \quad u_i(x_i) \\ &\text{s.t. } p^* \cdot x_i \le p^* \cdot \omega_i + \sum_{j=1}^{J} \theta_{ij}(p^* \cdot y_j^*). \end{aligned} \tag{10.B.2}$$

(iii) *Market clearing*: For each good $\ell = 1, \ldots, L$,

$$\sum_{i=1}^{I} x_{\ell i}^* = \omega_\ell + \sum_{j=1}^{J} y_{\ell j}^*. \tag{10.B.3}$$

Definition 10.B.3 delineates three sorts of conditions that must be met for a competitive economy to be considered to be in equilibrium. Conditions (i) and (ii) reflect the underlying assumption, common to nearly all economic models, that agents in the economy seek to do as well as they can for themselves. Condition (i) states that each firm must choose a production plan that maximizes its profits, taking as given the equilibrium vector of prices of its outputs and inputs (for the justification of the profit-maximization assumption, see Section 5.G). We studied this competitive behavior of the firm extensively in Chapter 5.

Condition (ii) requires that each consumer chooses a consumption bundle that maximizes his utility given the budget constraint imposed by the equilibrium prices and by his wealth. We studied this competitive behavior of the consumer extensively in Chapter 3. One difference here, however, is that the consumer's wealth is now a function of prices. This dependence of wealth on prices arises in

1. Strictly speaking, it is *equilibrium* market prices that they will regard as unaffected by their actions. For more on this point, see the small-type discussion later in this section.

two ways: First, prices determine the value of the consumer's initial endowments; for example, an individual who initially owns real estate is poorer if the price of real estate falls. Second, the equilibrium prices affect firms' profits and hence the value of the consumer's shareholdings.

Condition (iii) is somewhat different. It requires that, at the equilibrium prices, the desired consumption and production levels identified in conditions (i) and (ii) are in fact mutually compatible; that is, the aggregate supply of each commodity (its total endowment plus its net production) equals the aggregate demand for it. If excess supply or demand existed for a good at the going prices, the economy could not be at a point of equilibrium. For example, if there is excess demand for a particular commodity at the existing prices, some consumer who is not receiving as much of the commodity as he desires could do better by offering to pay just slightly more than the going market price and thereby get sellers to offer the commodity to him first. Similarly, if there is excess supply, some seller will find it worthwhile to offer his product at a slight discount from the going market price.[2]

Note that in justifying why an equilibrium must involve no excess demand or supply, we have actually made use of the fact that consumers and producers *might not* simply take market prices as given. How are we to reconcile this argument with the underlying price-taking assumption?

An answer to this apparent paradox comes from recognizing that consumers and producers *always* have the ability to alter their offered prices (in the absence of any institutional constraints preventing this). For the price-taking assumption to be appropriate, what we want is that they have no *incentive* to alter prices that, if taken as given, equate demand and supply (we have already seen that they *do* have an incentive to alter prices that do not equate demand and supply).

Notice that as long as consumers can make their desired trades at the going market prices, they will not wish to offer more than the market price to entice sellers to sell to them. Similarly, if producers are able to make their desired sales, they will have no incentive to undercut the market price. Thus, at a price that equates demand and supply, consumers do not wish to raise prices, and firms do not wish to lower them.

More troublesome is the possibility that a buyer might try to lower the price he pays or that a seller might try to raise the price he charges. A seller, for example, may possess the ability to raise profitably prices of the goods he sells above their competitive level (see Chapter 12). In this case, there is no reason to believe that this market power will not be exercised. To rescue the price-taking assumption, one needs to argue that under appropriate (competitive) conditions such market power does not exist. This we do in Sections 12.F and 18.C, where we formalize the idea that if market participants' desired trades are small relative to the size of the market, then they will have little incentive to depart from market prices. Thus, in a suitably defined equilibrium, they will act approximately like price takers.

Note from Definition 10.B.3 that if the allocation $(x_1^*, \ldots, x_I^*, y_1^*, \ldots, y_J^*)$ and price vector $p^* \gg 0$ constitute a competitive equilibrium, then so do the allocation

2. Strictly speaking, this second part of the argument requires the price to be positive; indeed, if the price is zero (i.e., if the good is free), then excess supply should be permissible at equilibrium. In the remainder of this chapter, however, consumer preferences will be such as to preclude this possibility (goods will be assumed to be desirable). Hence, we neglect this possibility here.

$(x_1^*, \ldots, x_I^*, y_1^*, \ldots, y_J^*)$ and price vector $\alpha p^* = (\alpha p_1^*, \ldots, \alpha p_L^*)$ for any scalar $\alpha > 0$ (see Exercise 10.B.2). As a result, we can normalize prices without loss of generality. In this chapter, we always normalize by setting one good's price equal to 1.

Lemma 10.B.1 will also prove useful in identifying competitive equilibria.

Lemma 10.B.1: If the allocation $(x_1, \ldots, x_I, y_1, \ldots, y_J)$ and price vector $p \gg 0$ satisfy the market clearing condition (10.B.3) for all goods $\ell \neq k$, and if every consumer's budget constraint is satisfied with equality, so that $p \cdot x_i = p \cdot \omega_i + \sum_j \theta_{ij} p \cdot y_j$ for all i, then the market for good k also clears.

Proof: Adding up the consumers' budget constraints over the I consumers and rearranging terms, we get

$$\sum_{\ell \neq k} p_\ell \left(\sum_{i=1}^{I} x_{\ell i} - \omega_\ell - \sum_{j=1}^{J} y_{\ell j} \right) = -p_k \left(\sum_{i=1}^{I} x_{ki} - \omega_k - \sum_{j=1}^{J} y_{kj} \right).$$

By market clearing in goods $\ell \neq k$, the left-hand side of this equation is equal to zero. Thus, the right-hand side must be equal to zero as well. Because $p_k > 0$, this implies that we have market clearing in good k. ■

In the models studied in this chapter, Lemma 10.B.1 will allow us to identify competitive equilibria by checking for market clearing in only $L - 1$ markets. Lemma 10.B.1 is really just a matter of double-entry accountancy. If consumers' budget constraints hold with equality, the dollar value of each consumer's planned purchases equals the dollar value of what he plans to sell plus the dollar value of his share (θ_{ij}) of the firms' (net) supply, and so the total value of planned purchases in the economy must equal the total value of planned sales. If those values are equal to each other in all markets but one, then equality must hold in the remaining market as well.

10.C Partial Equilibrium Competitive Analysis

Marshallian partial equilibrium analysis envisions the market for one good (or several goods, as discussed in Section 10.G) that constitutes a small part of the overall economy. The small size of the market facilitates two important simplifications for the analysis of market equilibrium:[3] First, as Marshall (1920) emphasized, when the expenditure on the good under study is a small portion of a consumer's total expenditure, only a small fraction of any additional dollar of wealth will be spent on this good; consequently, we can expect wealth effects for it to be small. Second, with similarly dispersed substitution effects, the small size of the market under study should lead the prices of other goods to be approximately unaffected by changes in this market.[4] Because of this fixity of other prices, we are justified in treating the expenditure on these other goods as a single composite commodity, which we call the *numeraire* (see Exercise 3.G.5).

3. The following points have been formalized by Vives (1987). (See Exercise 10.C.1 for an illustration.)

4. This is not the only possible justification for taking other goods' prices as being unaffected by the market under study; see Section 10.G.

With this partial equilibrium interpretation as our motivation, we proceed to study a simple two-good quasilinear model. There are two commodities: good ℓ and the numeraire. We let x_i and m_i denote consumer i's consumption of good ℓ and the numeraire, respectively. Each consumer $i = 1, \ldots, I$ has a utility function that takes the quasilinear form (see Sections 3.B and 3.C):

$$u_i(m_i, x_i) = m_i + \phi_i(x_i).$$

We let each consumer's consumption set be $\mathbb{R} \times \mathbb{R}_+$, and so we assume for convenience that consumption of the numeraire commodity m can take negative values. This is to avoid dealing with boundary problems. We assume that $\phi_i(\cdot)$ is bounded above and twice differentiable, with $\phi_i'(x_i) > 0$ and $\phi_i''(x_i) < 0$ at all $x_i \geq 0$. We normalize $\phi_i(0) = 0$.

In terms of our partial equilibrium interpretation, we think of good ℓ as the good whose market is under study and of the numeraire as representing the composite of all other goods (m stands for the total money expenditure on these other goods). Recall that with quasilinear utility functions, wealth effects for non-numeraire commodities are null.

In the discussion that follows, we normalize the price of the numeraire to equal 1, and we let p denote the price of good ℓ.

Each firm $j = 1, \ldots, J$ in this two-good economy is able to produce good ℓ from good m. The amount of the numeraire required by firm j to produce $q_j \geq 0$ units of good ℓ is given by the cost function $c_j(q_j)$ (recall that the price of the numeraire is 1). Letting z_j denote firm j's use of good m as an input, its production set is therefore

$$Y_j = \{(-z_j, q_j)\colon\ q_j \geq 0 \text{ and } z_j \geq c_j(q_j)\}.$$

In what follows, we assume that $c_j(\cdot)$ is twice differentiable, with $c_j'(q_j) > 0$ and $c_j''(q_j) \geq 0$ at all $q_j \geq 0$. [In terms of our partial equilibrium interpretation, we can think of $c_j(q_j)$ as actually arising from some multiple-input cost function $c_j(\bar{w}, q_j)$, given the fixed vector of factor prices $\bar{w}$.[5]]

For simplicity, we shall assume that there is no initial endowment of good ℓ, so that all amounts consumed must be produced by the firms. Consumer i's initial endowment of the numeraire is the scalar $\omega_{mi} > 0$, and we let $\omega_m = \sum_i \omega_{mi}$.

We now proceed to identify the competitive equilibria for this two-good quasilinear model. Applying Definition 10.B.3, we consider first the implications of profit and utility maximization.

Given the price p^* for good ℓ, firm j's equilibrium output level q_j^* must solve

$$\operatorname*{Max}_{q_j \geq 0} \quad p^* q_j - c_j(q_j),$$

which has the necessary and sufficient first-order condition

$$p^* \leq c_j'(q_j^*), \quad \text{with equality if } q_j^* > 0.$$

On the other hand, consumer i's equilibrium consumption vector (m_i^*, x_i^*) must

5. Some of the exercises at the end of the chapter investigate the effects of exogenous changes in these factor prices.

solve

$$\underset{m_i \in \mathbb{R},\, x_i \in \mathbb{R}_+}{\text{Max}} \quad m_i + \phi_i(x_i)$$

$$\text{s.t.} \quad m_i + p^* x_i \leq \omega_{mi} + \sum_{j=1}^{J} \theta_{ij}(p^* q_j^* - c_j(q_j^*)).$$

In any solution to this problem, the budget constraint holds with equality. Substituting for m_i from this constraint, we can rewrite consumer i's problem solely in terms of choosing his optimal consumption of good ℓ. Doing so, we see that x_i^* must solve

$$\underset{x_i \geq 0}{\text{Max}} \quad \phi_i(x_i) - p^* x_i + \left[\omega_{mi} + \sum_{j=1}^{J} \theta_{ij}(p^* q_j^* - c_j(q_j^*))\right],$$

which has the necessary and sufficient first-order condition

$$\phi_i'(x_i^*) \leq p^*, \quad \text{with equality if } x_i^* > 0.$$

In what follows, it will be convenient to adopt the convention of identifying an equilibrium allocation by the levels of good ℓ consumed and produced, $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$, with the understanding that consumer i's equilibrium consumption of the numeraire is then $m_i^* = [\omega_{mi} + \sum_j \theta_{ij}(p^* q_j^* - c_j(q_j^*))] - p^* x_i^*$ and that firm j's equilibrium usage of the numeraire as an input is $z_j^* = c_j(q_j^*)$.

To complete the development of the equilibrium conditions for this model, recall that by Lemma 10.B.1, we need only check that the market for good ℓ clears.[6] Hence, we conclude that the allocation $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ and the price p^* constitute a competitive equilibrium if and only if

$$p^* \leq c_j'(q_j^*), \quad \text{with equality if } q_j^* > 0 \qquad j = 1, \ldots, J. \tag{10.C.1}$$

$$\phi_i'(x_i^*) \leq p^*, \quad \text{with equality if } x_i^* > 0 \qquad i = 1, \ldots, I. \tag{10.C.2}$$

$$\sum_{i=1}^{I} x_i^* = \sum_{j=1}^{J} q_j^*. \tag{10.C.3}$$

At any interior solution, condition (10.C.1) says that firm j's marginal benefit from selling an additional unit of good ℓ, p^*, exactly equals its marginal cost $c_j'(q_j^*)$. Condition (10.C.2) says that consumer i's marginal benefit from consuming an additional unit of good ℓ, $\phi_i'(x_i^*)$, exactly equals its marginal cost p^*. Condition (10.C.3) is the market-clearing equation. Together, these $I + J + 1$ conditions characterize the $(I + J + 1)$ equilibrium values $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ and p^*. Note that as long as $\text{Max}_i\, \phi_i'(0) > \text{Min}_j\, c_j'(0)$, the aggregate consumption and production of good ℓ must be strictly positive in a competitive equilibrium [this follows from conditions (10.C.1) and (10.C.2)]. For simplicity, we assume that this is the case in the discussion that follows.

Conditions (10.C.1) to (10.C.3) have a very important property: They do not involve, in any manner, the endowments or the ownership shares of the consumers. As a result, we see that *the equilibrium allocation and price are independent of the*

6. Note that we must have $p^* > 0$ in any competitive equilibrium; otherwise, consumers would demand an infinite amount of good ℓ [recall that $\phi_i'(\cdot) > 0$].

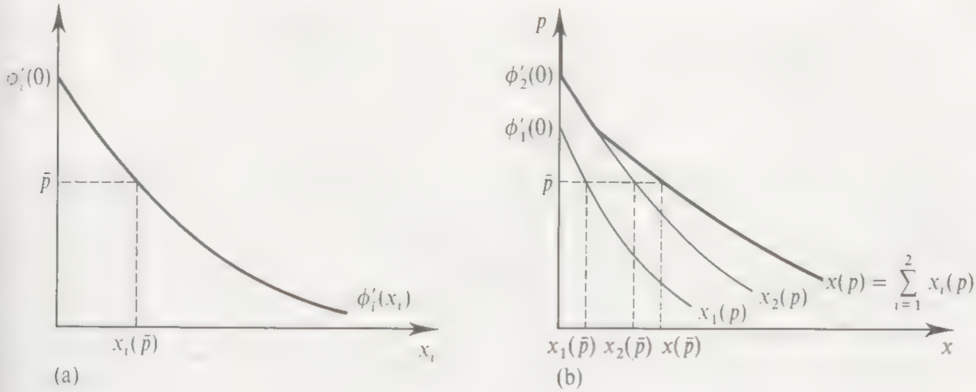


Figure 10.C.1 Construction of the aggregate demand function. (a) Determination of consumer i's demand. (b) Construction of the aggregate demand function ($I = 2$).

distribution of endowments and ownership shares. This important simplification arises from the quasilinear form of consumer preferences.[7]

The competitive equilibrium of this model can be nicely represented using the traditional Marshallian graphical technique that identifies the equilibrium price as the point of intersection of aggregate demand and aggregate supply curves.

We can derive the aggregate demand function for good ℓ from condition (10.C.2). Because $\phi_i''(\cdot) < 0$ and $\phi_i(\cdot)$ is bounded, $\phi_i'(\cdot)$ is a strictly decreasing function of x_i taking all values in the set $(0, \phi_i'(0)]$. Therefore, for each possible level of $p > 0$, we can solve for a unique level of x_i, denoted $x_i(p)$, that satisfies condition (10.C.2). Note that if $p \geq \phi_i'(0)$, then $x_i(p) = 0$. Figure 10.C.1(a) depicts this construction for a price $\bar{p} > 0$. The function $x_i(\cdot)$ is consumer i's *Walrasian demand function* for good ℓ (see Section 3.D) which, because of quasilinearity, does not depend on the consumer's wealth. It is continuous and nonincreasing in p at all $p > 0$, and is strictly decreasing at any $p < \phi_i'(0)$ [at any such p, we have $x_i'(p) = 1/\phi_i''(x_i(p)) < 0$].

The *aggregate demand function for good ℓ* is then the function $x(p) = \sum_i x_i(p)$, which is continuous and nonincreasing at all $p > 0$, and is strictly decreasing at any $p < \text{Max}_i\, \phi_i'(0)$. Its construction is depicted in Figure 10.C.1(b) for the case in which $I = 2$; it is simply the horizontal summation of the individual demand functions and is drawn in the figure with a heavy trace. Note that $x(p) = 0$ whenever $p \geq \text{Max}_i\, \phi_i'(0)$.

The aggregate supply function can be similarly derived from condition (10.C.1).[8] Suppose, first, that every $c_j(\cdot)$ is strictly convex and that $c_j'(q_j) \to \infty$ as $q_j \to \infty$. Then, for any $p > 0$, we can let $q_j(p)$ denote the unique level of q_j that satisfies condition (10.C.1). Note that for $p \leq c_j'(0)$, we have $q_j(p) = 0$. Figure 10.C.2(a) illustrates this construction for a price $\bar{p} > 0$. The function $q_j(\cdot)$ is firm j's *supply function* for good ℓ (see Sections 5.C and 5.D). It is continuous and nondecreasing at all $p > 0$, and is strictly increasing at any $p > c_j'(0)$ [for any such p, $q_j'(p) = 1/c_j''(q_j(p)) > 0$].

The *aggregate* (or *industry*) *supply function* for good ℓ is then the function $q(p) = \sum_j q_j(p)$, which is continuous and nondecreasing at all $p > 0$, and is strictly increasing at any $p > \text{Min}_j\, c_j'(0)$. Its construction is depicted in Figure 10.C.2(b) for

[7] See Section 10.G for a further discussion of this general feature of equilibrium in economies with quasilinear utility functions.

[8] See Section 5.D for an extensive discussion of individual supply in the one-input, one-output case.

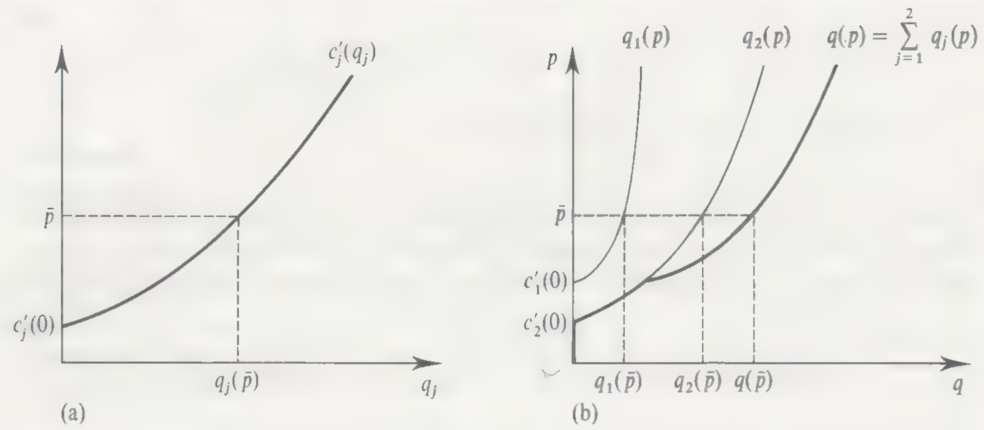


Figure 10.C.2 Construct... aggregate ... function. (a) Dete... firm j's ... (b) Con... the aggr... function (J ...

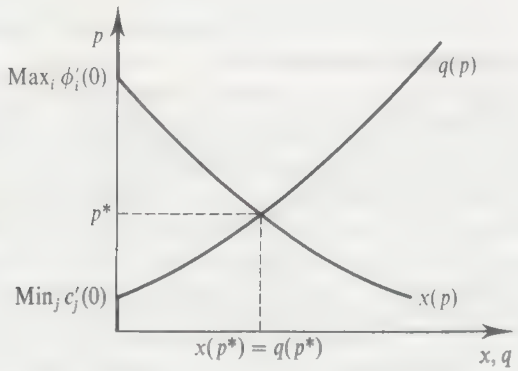


Figure 10.C.3 The equi... equates de... supply.

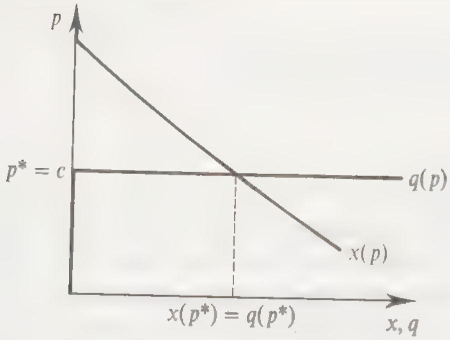
the case in which $J = 2$; it is equal to the horizontal sum of the individual firms' supply functions and is drawn in the figure with a heavy trace. Note that $q(p) = 0$ whenever $p \leq \text{Min}_j\, c_j'(0)$.

To find the equilibrium price of good ℓ, we need only find the price p^* at which aggregate demand equals aggregate supply, that is, at which $x(p^*) = q(p^*)$. When $\text{Max}_i\, \phi_i'(0) > \text{Min}_j\, c_j'(0)$ as we have assumed, at any $p \geq \text{Max}_i\, \phi_i'(0)$ we have $x(p) = 0$ and $q(p) > 0$. Likewise, at any $p \leq \text{Min}_j\, c_j'(0)$ we have $x(p) > 0$ and $q(p) = 0$. The existence of an equilibrium price $p^* \in (\text{Min}_j\, c_j'(0), \text{Max}_i\, \phi_i'(0))$ then follows from the continuity properties of $x(\cdot)$ and $q(\cdot)$. The solution is depicted in Figure 10.C.3. Note also that because $x(\cdot)$ is strictly decreasing at all $p < \text{Max}_i\, \phi_i'(0)$ and $q(\cdot)$ is strictly increasing at all $p > \text{Min}_j\, c_j'(0)$, this equilibrium price is uniquely defined.[9] The individual consumption and production levels of good ℓ in this equilibrium are then given by $x_i^* = x_i(p^*)$ for $i = 1, \ldots, I$ and $q_j^* = q_j(p^*)$ for $j = 1, \ldots, J$.

More generally, if some $c_j(\cdot)$ is merely convex [e.g., if $c_j(\cdot)$ is linear, as in the constant returns case], then $q_j(\cdot)$ is a convex-valued correspondence rather than a function and it may be well defined only on a subset of prices.[10] Nevertheless, the

9. Be warned, however, that the uniqueness of equilibrium is a property that need not hold in more general settings in which wealth effects are present. (See Chapter 17.)

10. For example, if firm j has $c_j(q_j) = c_j q_j$ for some scalar $c_j > 0$, then when $p > c_j$, we have $q_j(p) = \infty$. As a result, if $p > c_j$, the aggregate supply is $q(p) = \sum_j q_j(p) = \infty$; consequently $q(\cdot)$ is not well defined for this p.



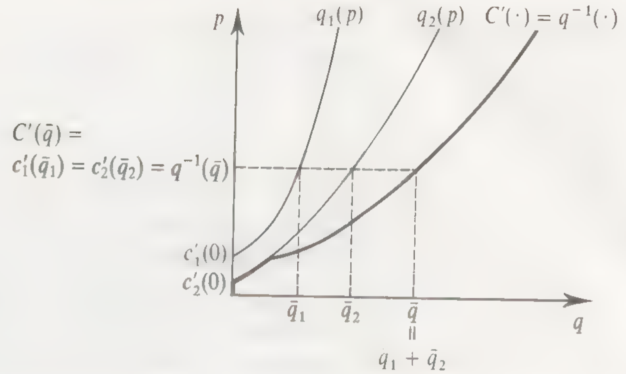


Figure 10.C.4 (left) Equilibrium when $c_j(q_j) = cq_j$ for all $j = 1, \ldots, J$.

Figure 10.C.5 (right) The industry marginal cost function.

basic features of the analysis do not change. Figure 10.C.4 depicts the determination of the equilibrium value of p in the case where, for all j, $c_j(q_j) = cq_j$ for some scalar $c > 0$. The only difference from the strictly convex case is that, when $J > 1$, individual firms' equilibrium production levels are not uniquely determined.

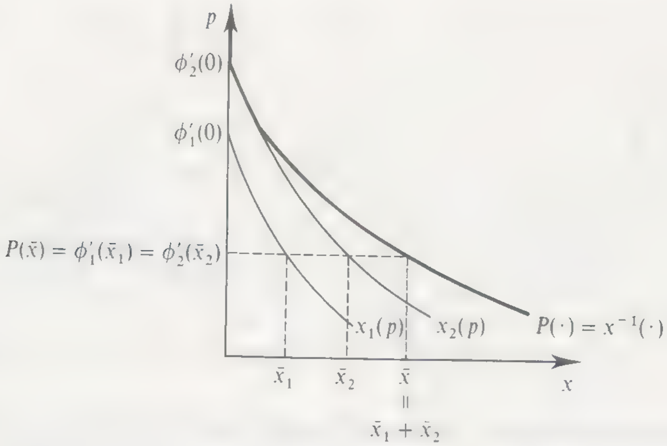
The inverses of the aggregate demand and supply functions also have interpretations that are of interest. At any given level of aggregate output of good ℓ, say $\bar{q}$, the inverse of the industry supply function, $q^{-1}(\bar{q})$, gives the price that brings forth aggregate supply $\bar{q}$. That is, when each firm chooses its optimal output level facing the price $p = q^{-1}(\bar{q})$, aggregate supply is exactly $\bar{q}$. Figure 10.C.5 illustrates this point. Note that in selecting these output levels, all active firms set their marginal cost equal to $q^{-1}(\bar{q})$. As a result, the marginal cost of producing an additional unit of good ℓ at $\bar{q}$ is precisely $q^{-1}(\bar{q})$, regardless of which active firm produces it. Thus $q^{-1}(\cdot)$, the inverse of the industry supply function, can be viewed as the *industry marginal cost function*, which we now denote by $C'(\cdot) = q^{-1}(\cdot)$.[11]

The derivation of $C'(\cdot)$ just given accords fully with our discussion in Section 5.E. We saw there that the aggregate supply of the J firms, $q(p)$, maximizes aggregate profits given p; therefore, we can relate $q(\cdot)$ to the industry marginal cost function $C'(\cdot)$ in exactly the same manner as we did in Section 5.D for the case of a single firm's marginal cost function and supply behavior. With convex technologies, the aggregate supply locus for good ℓ therefore coincides with the graph of the industry marginal cost function $C'(\cdot)$, and so $q^{-1}(\cdot) = C'(\cdot)$.[12]

Likewise, at any given level of aggregate demand $\bar{x}$, the *inverse demand function* $P(\bar{x}) = x^{-1}(\bar{x})$ gives the price that results in aggregate demand of $\bar{x}$. That is, when each consumer optimally chooses his demand for good ℓ at this price, total demand exactly equals $\bar{x}$. Note that at these individual demand levels (assuming that they are positive), each consumer's marginal benefit in terms of the numeraire from an additional unit of good ℓ, $\phi_i'(x_i)$, is exactly equal to $P(\bar{x})$. This is illustrated in Figure

11. Formally, the industry marginal cost function $C'(\cdot)$ is the derivative of the aggregate cost function $C(\cdot)$ that gives the total production cost that would be incurred by a central authority who operates all J firms and seeks to produce any given aggregate level of good ℓ at minimum total cost. (See Exercise 10.C.3.)

12. More formally, by Proposition 5.E.1, aggregate supply behavior can be determined by maximizing profit given the aggregate cost function $C(\cdot)$. This yields first-order condition $p = C'(q(p))$. Hence, $q(\cdot) = C'^{-1}(\cdot)$, or equivalently $q^{-1}(\cdot) = C'(\cdot)$.



10.C.6. The value of the inverse demand function at quantity $\bar{x}$, $P(\bar{x})$, can thus be viewed as giving the *marginal social benefit of good* ℓ given that the aggregate quantity $\bar{x}$ is efficiently distributed among the I consumers (see Exercise 10.C.4 for a precise statement of this fact).

Given these interpretations, we can view the competitive equilibrium as involving an aggregate output level at which the marginal social benefit of good ℓ is exactly equal to its marginal cost. This suggests a social optimality property of the competitive allocation, a topic that we investigate further in Section 10.D.

Comparative Statics

It is often of interest to determine how a change in underlying market conditions affects the equilibrium outcome of a competitive market. Such questions may arise, for example, because we may be interested in comparing market outcomes across several similar markets that differ in some measurable way (e.g., we might compare the price of ice cream in a number of cities whose average temperatures differ) or because we want to know how a change in market conditions will alter the outcome in a particular market. The analysis of these sorts of questions is known as *comparative statics analysis.*

As a general matter, we might imagine that each consumer's preferences are affected by a vector of exogenous parameters $\alpha \in \mathbb{R}^M$, so that the utility function $\phi_i(\cdot)$ can be written as $\phi_i(x_i, \alpha)$. Similarly, each firm's technology may be affected by a vector of exogenous parameters $\beta \in \mathbb{R}^S$, so that the cost function $c_j(\cdot)$ can be written as $c_j(q_j, \beta)$. In addition, in some circumstances, consumers and firms face taxes or subsidies that may make the effective (i.e., net of taxes and subsidies) price paid or received differ from the market price p. We let $\hat{p}_i(p, t)$ and $\hat{p}_j(p, t)$ denote, respectively, the effective price paid by consumer i and the effective price received by firm j given tax and subsidy parameters $t \in \mathbb{R}^K$. For example, if consumer i must pay a tax of t_i (in units of the numeraire) per unit of good i purchased, then $\hat{p}_i(p, t) = p + t_i$. If consumer i instead faces a tax that is a percentage t_i of the sales price, then $\hat{p}_i(p, t) = p(1 + t_i)$.

For given values (α, β, t) of the parameters, the $I + J$ equilibrium quantities $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ and the equilibrium price p^* are determined as the solution to the following $I + J + 1$ equations (we assume, for simplicity, that $x_i^* > 0$ for all

and $q_j^* > 0$ for all j):

$$\phi_i'(x_i^*, \alpha) = \hat{p}_i(p^*, t) \qquad i = 1, \ldots, I. \tag{10.C.4}$$

$$c_j'(q_j^*, \beta) = \hat{p}_j(p^*, t) \qquad j = 1, \ldots, J. \tag{10.C.5}$$

$$\sum_{i=1}^{I} x_i^* = \sum_{j=1}^{J} q_j^*. \tag{10.C.6}$$

These $I + J + 1$ equations implicitly define the equilibrium allocation and price as ...ns of the exogenous parameters (α, β, t). If all the relevant functions are ...entiable, we can use the implicit function theorem to derive the marginal change ... equilibrium allocation and price in response to a differential change in the ... of these parameters (see Section M.E of the Mathematical Appendix). In ...ple 10.C.1, we consider one such comparative statics exercise; it is only one ... a large number of possibilities that arise naturally in economic applications. ...ercises at the end of this chapter include additional examples.)

...ple **10.C.1:** *Comparative Statics Effects of a Sales Tax.* Suppose that a new sales ... proposed under which consumers must pay an amount $t \geq 0$ (in units of the ...re) for each unit of good ℓ consumed. We wish to determine the effect of ... on the market price. Let $x(p)$ and $q(p)$ denote the aggregate demand and ... functions, respectively, for good ℓ in the absence of the tax (we maintain all ... previous assumptions regarding these functions).

...rms of our previous notation, the $\phi_i(\cdot)$ and $c_j(\cdot)$ functions do not depend ... exogenous parameters, $\hat{p}_i(p, t) = p + t$ for all i, and $\hat{p}_j(p, t) = p$ for all j. In ... by substituting these expressions into the system of equilbrium equations ... to (10.C.6), we can derive the effect of a marginal increase in the tax on the ... direct use of the implicit function theorem (see Exercise 10.C.5). Here, ... we pursue a more instructive way to get the answer. In particular, note ...egate demand with a tax of t and price p is exactly $x(p + t)$ because the tax ...ent for consumers to the price being increased by t. Thus, the equilibrium ... price when the tax is t, which we denote by $p^*(t)$, must satisfy

$$x(p^*(t) + t) = q(p^*(t)). \tag{10.C.7}$$

...se that we now want to determine the effect on prices paid and received ...inal increase in the tax. Assuming that $x(\cdot)$ and $q(\cdot)$ are differentiable at ...), differentiating (10.C.7) yields

$$p^{*\prime}(t) = -\frac{x'(p^*(t) + t)}{x'(p^*(t) + t) - q'(p^*(t))}. \tag{10.C.8}$$

...ediate from (10.C.8) and our assumptions on $x'(\cdot)$ and $q'(\cdot)$ that $-1 \leq$ ... at any t. Therefore, the price $p^*(t)$ received by producers falls as t increases ... overall cost of the good to consumers $p^*(t) + t$ rises (weakly). The total ... produced and consumed fall (again weakly). See Figure 10.C.7(a), where ...rium level of aggregate consumption at tax rate t is denoted by $x^*(t)$. ... (10.C.8) that when $q'(p^*(t))$ is large we have $p^{*\prime}(t) \approx 0$, and so the price ... the firms is hardly affected by the tax; nearly all the impact of the tax is ...umers. In contrast, when $q'(p^*(t)) = 0$, we have $p^{*\prime}(t) = -1$, and so the ... the tax is felt entirely by the firms. Figures 10.C.7(b) and (c) depict these

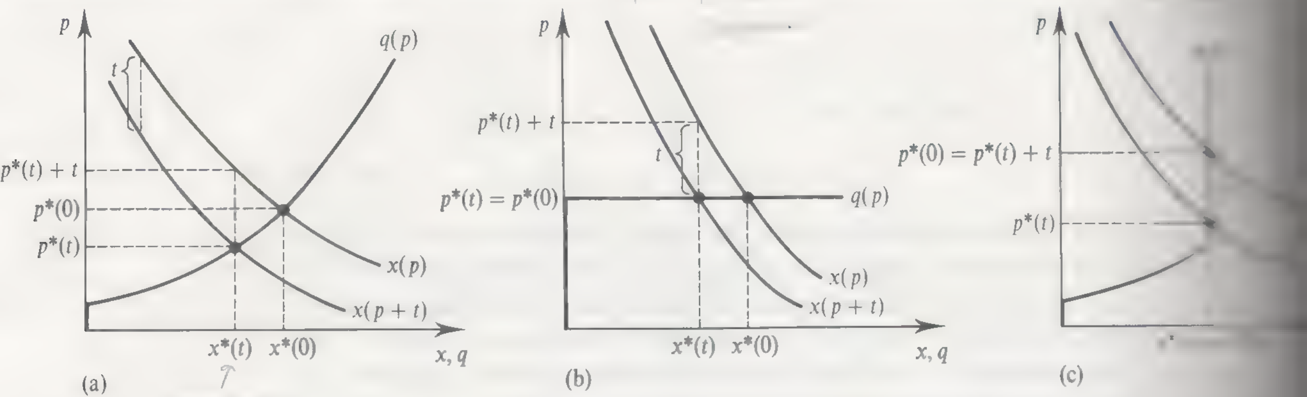


Figure 10.C.7 Comparative statics effects of a sales tax.

By substituting into (10.C.8) for $x'(\cdot)$ and $q'(\cdot)$, the marginal change in p^* can be expressed in terms of derivatives of the underlying individual utility and cost functions. For example, if we let $p^* = p^*(0)$ be the pretax price, we see that

$$p^{*\prime}(0) = -\frac{\sum_{i=1}^{I}[\phi_i''(x_i(p^*))]^{-1}}{\sum_{i=1}^{I}[\phi_i''(x_i(p^*))]^{-1} - \sum_{j=1}^{J}[c_j''(q_j(p^*))]^{-1}}.$$

■

We have assumed throughout this section that consumers' preferences and firms' technologies are convex (and strictly so in the case of consumer preferences). What if this is not the case? Figure 10.C.8 illustrates one problem that can then arise; it shows the demand function and

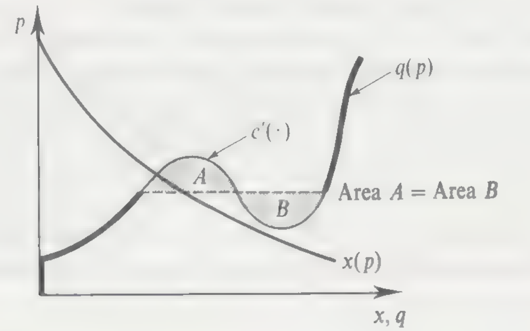


Figure ...
None...
compe...
equil...
nonc...

supply correspondence for an economy in which there is a single firm (so $J = 1$).[13] This firm's cost function $c(\cdot)$ is continuous and differentiable but not convex. In the figure, the light curve is the graph of the firm's marginal cost function $c'(\cdot)$. As the figure illustrates, $c'(\cdot)$ fails to be nondecreasing. The heavier curve is the firm's actual supply correspondence $q(\cdot)$ (you should verify that it is determined as indicated in the figure).[14] The graph of the supply correspondence no longer coincides with the marginal cost curve and, as is evident in the figure, no intersection exists between the graph of the supply correspondence and the demand curve. Thus, in this case, *no competitive equilibrium exists.*

13. We set $J = 1$ here solely for expositional purposes.

14. See Section 5.D for a more detailed discussion of the relation between a firm's supply correspondence and its marginal cost function when its technology is nonconvex.

This observation suggests that convexity assumptions are key to the existence of a competitive equilibrium. We shall confirm this in Chapter 17, where we provide a more general discussion of the conditions under which existence of a competitive equilibrium is assured.

The Fundamental Welfare Theorems in a Partial Equilibrium Context

In this section, we study the properties of Pareto optimal allocations in the framework of the two-good quasilinear economy introduced in Section 10.C, and we establish a fundamental link between the set of Pareto optimal allocations and the set of competitive equilibria.

The identification of Pareto optimal allocations is considerably facilitated by the quasilinear specification. In particular, *when consumer preferences are quasilinear, the boundary of the economy's utility possibility set is linear* (see Section 10.B for the definition of this set) *and all points in this boundary are associated with consumption allocations that differ only in the distribution of the numeraire among consumers.*

To see this important fact, suppose that we fix the consumption and production levels of good ℓ at $(\bar{x}_1, \ldots, \bar{x}_I, \bar{q}_1, \ldots, \bar{q}_J)$. With these production levels, the total amount of the numeraire available for distribution among consumers is $\omega_m - \sum_j c_j(\bar{q}_j)$. Because the quasilinear form of the utility functions allows for an unlimited unit-for-unit transfer of utility across consumers through transfers of the numeraire, the set of utilities that can be attained for the I consumers by appropriately distributing the available amounts of the numeraire is given by

$$\left\{(u_1, \ldots, u_I) \colon \sum_{i=1}^{I} u_i \le \sum_{i=1}^{I} \phi_i(\bar{x}_i) + \omega_m - \sum_{j=1}^{J} c_j(\bar{q}_j)\right\}. \qquad \text{(10.D.1)}$$

The boundary of this set is a hyperplane with normal vector $(1, \ldots, 1)$. The set is depicted for the case $I = 2$ by the hatched set in Figure 10.D.1.

Note that by altering the consumption and production levels of good ℓ, we necessarily shift the boundary of this set in a parallel manner. Thus, every Pareto optimal allocation must involve the quantities $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ that extend the boundary as far out as possible, as illustrated by the heavily drawn boundary of the shaded utility possibility set in Figure 10.D.1. We call these quantities the *optimal consumption and production levels* for good ℓ. As long as these optimal consumption and production levels for good ℓ are uniquely determined, Pareto optimal allocations can differ only in the distribution of the numeraire among consumers.[15]

15. The optimal individual production levels need not be unique if firms' cost functions are convex but not strictly so. Indeterminacy of optimal individual production levels arises, for example, when all firms have identical constant returns to scale technologies. However, under our assumptions that the $\phi_i(\cdot)$ functions are strictly concave and that the $c_j(\cdot)$ functions are convex, the optimal individual consumption levels of good ℓ are necessarily unique and, hence, so is the optimal aggregate production level $\sum_j q_j^*$ of good ℓ. This implies that, under our assumptions, the consumption levels in two different Pareto optimal allocations can differ only in the distribution of numeraire among consumers. If, moreover, the $c_j(\cdot)$ functions are strictly convex, then the optimal individual production levels are also uniquely determined. (See Exercise 10.D.1.)

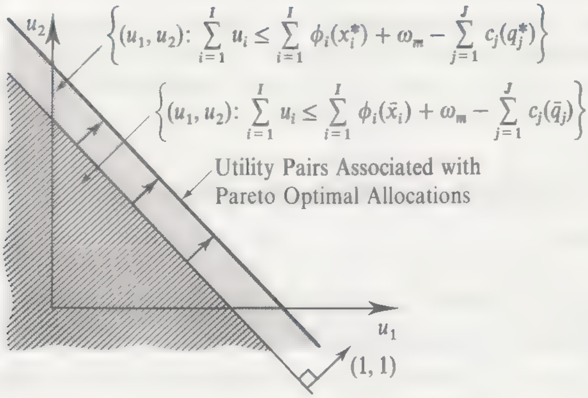


Figure [illegible]
The [illegible] set [illegible] econ[illegible]

It follows from expression (10.D.1) that the optimal consumption and production levels of good ℓ can be obtained as the solution to

$$\begin{aligned} \underset{\substack{(x_1,\ldots,x_I)\ge 0\\ (q_1,\ldots,q_J)\ge 0}}{\text{Max}} \quad & \sum_{i=1}^{I} \phi_i(x_i) - \sum_{j=1}^{J} c_j(q_j) + \omega_m \qquad (10.D.2)\\ \text{s.t.} \quad & \sum_{i=1}^{I} x_i - \sum_{j=1}^{J} q_j = 0. \end{aligned}$$

The value of the term $\sum_i \phi_i(x_i) - \sum_j c_j(q_j)$ in the objective function of problem (10.D.2) is known as the *Marshallian aggregate surplus* (or, simply, the *aggregate surplus*). It can be thought of as the total utility generated from consumption of good ℓ less its costs of production (in terms of the numeraire). The optimal consumption and production levels for good ℓ maximize this aggregate surplus measure.

Given our convexity assumptions, the first-order conditions of problem (10.D.2) yield necessary and sufficient conditions that characterize the optimal quantities. If we let μ be the multiplier on the constraint in problem (10.D.2), the $I + J$ optimal values $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ and the multiplier μ satisfy the following $I + J + 1$ conditions:

$$\mu \le c_j'(q_j^*), \quad \text{with equality if } q_j^* > 0 \qquad j = 1, \ldots, J. \tag{10.D.3}$$

$$\phi_i'(x_i^*) \le \mu, \quad \text{with equality if } x_i^* > 0 \qquad i = 1, \ldots, I. \tag{10.D.4}$$

$$\sum_{i=1}^{I} x_i^* = \sum_{j=1}^{J} q_j^*. \tag{10.D.5}$$

These conditions should look familiar: They exactly parallel conditions (10.C.1) to (10.C.3) in Section 10.C, with μ replacing p^*. This observation has an important implication. We can immediately infer from it that any competitive equilibrium outcome in this model is Pareto optimal because any competitive equilibrium allocation has consumption and production levels of good ℓ, $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$, that satisfy conditions (10.D.3) to (10.D.5) when we set $\mu = p^*$. Thus, we have established the *first fundamental theorem of welfare economics* (Proposition 10.D.1) in the context of this quasilinear two-good model.

Proposition 10.D.1: (*The First Fundamental Theorem of Welfare Economics*) If the price p^* and allocation $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$ constitute a competitive equilibrium, then this allocation is Pareto optimal.

The first fundamental welfare theorem establishes conditions under which market equilibria are necessarily Pareto optimal. It is a formal expression of Adam Smith's "invisible hand" and is a result that holds with considerable generality (see Section 16.C for a much more extensive discussion). Equally important, however, are the conditions under which it fails to hold. In the models for which we establish the first fundamental welfare theorem here and in Section 16.C, markets are "complete" in the sense that there is a market for every relevant commodity and all market participants act as price takers. In Chapters 11 to 14, we study situations in which at least one of these conditions fails, and market outcomes fail to be Pareto optimal as a result.

We can also develop a converse to Proposition 10.D.1, known as the *second fundamental theorem of welfare economics*. In Section 10.C, we saw that good ℓ's equilibrium price p^*, its equilibrium consumption and production levels $(x_1^*, \ldots, x_I^*, q_1^*, \ldots, q_J^*)$, and firms' profits are unaffected by changes in consumers' wealth levels. As a result, a transfer of one unit of the numeraire from consumer i to consumer i' will cause each of these consumers' equilibrium consumption of the numeraire to change by exactly the amount of the transfer and will cause no other changes. Thus, by appropriately transferring endowments of the numeraire commodity, the resulting competitive equilibrium allocation can be made to yield any utility vector on the boundary of the utility possibility set. The second welfare theorem therefore tells us that, in this two-good quasilinear economy, a central authority interested in achieving a particular Pareto optimal allocation can always implement this outcome by transferring the numeraire among consumers and then "allowing the market to work." This is stated formally in Proposition 10.D.2.

Proposition 10.D.2: (*The Second Fundamental Theorem of Welfare Economics*) For any Pareto optimal levels of utility $(u_1^*, \ldots, u_I^*)$, there are transfers of the numeraire commodity $(T_1, \ldots, T_I)$ satisfying $\sum_i T_i = 0$, such that a competitive equilibrium reached from the endowments $(\omega_{m1} + T_1, \ldots, \omega_{mI} + T_I)$ yields precisely the utilities $(u_1^*, \ldots, u_I^*)$.

In Section 16.D, we study the conditions under which the second welfare theorem holds in more general competitive economies. A critical requirement, in addition to those needed for the first welfare theorem, turns out to be convexity of preferences and production sets, an assumption we have made in the model under consideration here. In contrast, we shall see in Chapter 16 that no such convexity assumptions are needed for the first welfare theorem.

The correspondence between p and μ in the equilibrium conditions (10.C.1) to (10.C.3) and the Pareto optimality conditions (10.D.3) to (10.D.5) is worthy of emphasis: The competitive price is exactly equal to the shadow price on the resource constraint for good ℓ in the Pareto optimality problem (10.D.2). In this sense, then, we can say that a good's price in a competitive equilibrium reflects precisely its marginal social value. In a competitive equilibrium, each firm, by operating at a point where price equals marginal cost, equates its marginal production cost to the marginal social value of its output. Similarly, each consumer, by consuming up to the point where marginal utility from a good equals its price, is at a point where the marginal benefit from consumption of the good exactly equals its marginal cost. This correspondence between equilibrium market prices and optimal shadow prices holds

quite generally in competitive economies (see Section 16.F for further discussion of this point).

An alternative way to characterize the set of Pareto optimal allocations is to solve

$$\underset{\{x_i, m_i\}_{i=1}^{I}, \{z_j, q_j\}_{j=1}^{J}}{\text{Max}} \quad m_1 + \phi_1(x_1) \tag{10.D.6}$$

$$\begin{aligned} \text{s.t.} \quad &(1) \quad m_i + \phi_i(x_i) \geq \bar{u}_i \qquad i = 2, \ldots, I \\ &(2\ell) \quad \sum_{i=1}^{I} x_i - \sum_{j=1}^{J} q_j \leq 0 \\ &(2m) \quad \sum_{i=1}^{I} m_i + \sum_{j=1}^{J} z_j \leq \omega_m \\ &(3) \quad z_j \geq c_j(q_j) \qquad j = 1, \ldots, J. \end{aligned}$$

Problem (10.D.6) expresses the Pareto optimality problem as one of trying to maximize the well-being of individual 1 subject to meeting certain required utility levels for the other individuals in the economy [constraints (1)], resource constraints [constraints (2ℓ) and ($2m$)], and technological constraints [constraints (3)]. By solving problem (10.D.6) for various required levels of utility for these other individuals, $(\bar{u}_2, \ldots, \bar{u}_I)$, we can identify all the Pareto optimal outcomes for this economy (see Exercise 10.D.3; more generally, we can do this whenever consumer preferences are strongly monotone). Exercise 10.D.4 asks you to derive conditions (10.D.3) to (10.D.5) in this alternative manner.

10.E Welfare Analysis in the Partial Equilibrium Model

It is often of interest to measure the change in the level of social welfare that would be generated by a change in market conditions such as an improvement in technology, a new government tax policy, or the elimination of some existing market imperfection. In the partial equilibrium model, it is particularly simple to carry out this welfare analysis. This fact accounts to a large extent for the popularity of the model.

In the discussion that follows, we assume that the welfare judgments of society are embodied in a social welfare function $W(u_1, \ldots, u_I)$ assigning a social welfare value to every utility vector $(u_1, \ldots, u_I)$ (see Chapters 4, 16, and 22 for more on this concept). In addition, we suppose that (as in the theory of the normative representative consumer discussed in Section 4.D) there is some central authority who redistributes wealth by means of transfers of the numeraire commodity in order to maximize social welfare.[16] The critical simplification offered by the quasilinear specification of individual utility functions is that when there is a central authority who redistributes wealth in this manner, *changes in social welfare can be measured by changes in the Marshallian aggregate surplus* (introduced in Section 10.D) *for any social welfare function that society may have.*

To see this point (which we have in fact already examined in Example 4.D.2), consider some given consumption and production levels of good ℓ, $(x_1, \ldots, x_I, q_1, \ldots, q_J)$,

16. As in Section 4.D, we assume that consumers treat these transfers as independent of their own actions; that is, in the standard terminology, they are *lump-sum* transfers. You should think of the central authority as making the transfers prior to the opening of markets.

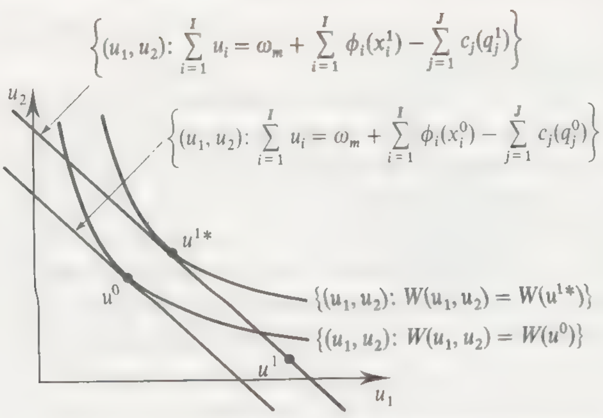


Figure 10.E.1 With lump-sum redistribution occurring to maximize social welfare, changes in welfare correspond to changes in aggregate surplus in a quasilinear model.

ng $\sum_i x_i = \sum_j q_j$. From Section 10.D and Figure 10.D.1 we know that the utility rs $(u_1, \ldots, u_I)$ that are achievable through reallocation of the numeraire given e consumption and production levels of good ℓ are

$$\left\{(u_1, \ldots, u_I)\colon \sum_{i=1}^{I} u_i \leq \omega_m + \sum_{i=1}^{I} \phi_i(x_i) - \sum_{j=1}^{J} c_j(q_j)\right\}.$$

if a central authority is redistributing the numeraire to maximize $W(u_1, \ldots, u_I)$, ltimate maximized value of welfare must be greater the larger this set is (i.e., rther out the boundary of the set is). Hence, we see that a change in the mption and production levels of good ℓ leads to an increase in welfare (given al redistribution of the numeraire) if and only if it increases the Marshallian regate surplus

$$S(x_1, \ldots, x_I, q_1, \ldots, q_J) = \sum_{i=1}^{I} \phi_i(x_i) - \sum_{j=1}^{J} c_j(q_j). \tag{10.E.1}$$

gure 10.E.1 provides an illustration. It shows three utility vectors for the case An initial utility vector $u^0 = (u_1^0, u_2^0)$ associated with an allocation in which nsumption and production levels of good ℓ are $(x_1^0, \ldots, x_I^0, q_1^0, \ldots, q_J^0)$ and in the wealth distribution has been optimized, a utility vector $u^1 = (u_1^1, u_2^1)$ that from a change in the consumption and production levels of good ℓ to $\ldots x_I^1, q_1^1, \ldots, q_J^1)$ in the absence of any transfers of the numeraire, and a utility $u^{1*} = (u_1^{1*}, u_2^{1*})$ that results from this change once redistribution of the aire occurs to optimize social welfare. As can be seen in the figure, the increases aggregate surplus and also increases welfare once optimal transfers numeraire occur, even though welfare would decrease in the absence of the . Thus, as long as redistribution of wealth is occurring to maximize a social function, changes in welfare can be measured by changes in Marshallian gate surplus (to repeat: for *any* social welfare function).[17]

many circumstances of interest, the Marshallian surplus has a convenient and

Notice that no transfers would be necessary in the special case in which the social welfare s in fact the "utilitarian" social welfare function $\sum_i u_i$; in this case, it is sufficient that all units of the numeraire go to consumers (i.e., none goes to waste or is otherwise withheld).

historically important formulation in terms of areas lying vertically between the aggregate demand and supply functions for good ℓ.

To expand on this point, we begin by making two key assumptions. Denoting by $x = \sum_i x_i$ the aggregate consumption of good ℓ, we assume, first, that for any x, the individual consumptions of good ℓ are distributed optimally across consumers. That is, recalling our discussion of the inverse demand function $P(\cdot)$ in Section 10.C (see Figure 10.C.6), that we have $\phi_i'(x_i) = P(x)$ for every i. This condition will be satisfied if, for example, consumers act as price-takers and all consumers face the same price. Similarly, denoting by $q = \sum_j q_j$ the aggregate output of good ℓ, we assume that the production of any total amount q is distributed optimally across firms. That is, recalling our discussion of the industry marginal cost curve $C'(\cdot)$ in Section 10.C (see Figure 10.C.5), that we have $c_j'(q_j) = C'(q)$ for every j. This will be satisfied if, for example, firms act as price takers and all firms face the same price. Observe that we do not require that the price faced by consumers and firms be the same.[18]

Consider now a differential change $(dx_1, \ldots, dx_I, dq_1, \ldots, dq_J)$ in the quantities of good ℓ consumed and produced satisfying $\sum_i dx_i = \sum_j dq_j$, and denote $dx = \sum_i dx_i$. The change in aggregate Marshallian surplus is then

$$dS = \sum_{i=1}^{I} \phi_i'(x_i)\, dx_i - \sum_{j=1}^{J} c_j'(q_j)\, dq_j. \tag{10.E.2}$$

Since $\phi_i'(x_i) = P(x)$ for all i, and $c_j'(q_j) = C'(q)$ for all j, we get

$$dS = P(x) \sum_{i=1}^{I} dx_i - C'(q) \sum_{j=1}^{J} dq_j. \tag{10.E.3}$$

Finally, since $x = q$ (by market feasibility) and $\sum_j dq_j = \sum_i dx_i = dx$, this becomes

$$dS = [P(x) - C'(x)]\, dx. \tag{10.E.4}$$

This differential change in Marshallian surplus is depicted in Figure 10.E.2(a). Expression (10.E.4) is quite intuitive; it tells us that starting at aggregate consumption level x the marginal effect on social welfare of an increase in the aggregate quantity consumed, dx, is equal to consumers' marginal benefit from this consumption, $P(x)\,dx$, less the marginal cost of this extra production, $C'(x)\,dx$ (both in terms of the numeraire).

We can also integrate (10.E.4) to express the total value of the aggregate Marshallian surplus at the aggregate consumption level x, denoted $S(x)$, in terms of an integral of the difference between the inverse demand function and the industry marginal cost function,

$$S(x) = S_0 + \int_0^x [P(s) - C'(s)]\, ds, \tag{10.E.5}$$

18. For example, consumers may face a tax per unit purchased that makes the price they pay differ from the price received by the firms (see Example 10.C.1). The assumptions made here also hold in the monopoly model to be studied in Section 12.B. In that model, there is a single firm (and so there is no issue of optimal allocation of production), and all consumers act as price takers facing the same price. An example where the assumption of an optimal allocation of production is not valid is the Cournot duopoly model of Chapter 12 when firms have different efficiencies. There, firms with different costs have different levels of marginal cost in an equilibrium.

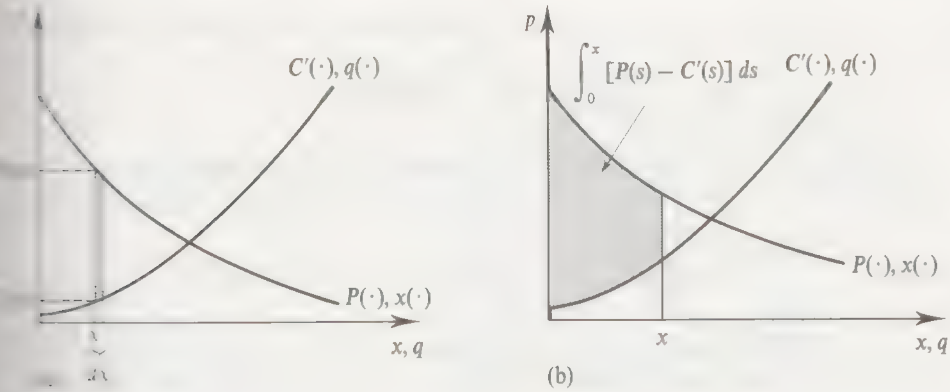


Figure 10.E.2
(a) A differential change in Marshallian surplus. (b) The Marshallian surplus at aggregate consumption level x.

... is a constant of integration equal to the value of the aggregate surplus when ... consumption or production of good ℓ [it is equal to zero if $c_j(0) = 0$ for ... The integral in (10.E.5) is depicted in Figure 10.E.2(b); it is exactly equal ... lying vertically between the aggregate demand and supply curves for good ... to quantity x.

... from (10.E.5) that the value of the aggregate Marshallian surplus is ...ed at the aggregate consumption level x^* such that $P(x^*) = C'(x^*)$, which ... the competitive equilibrium aggregate consumption level.[19] This accords ...position 10.D.1, the first fundamental welfare theorem, which states that the ...titive allocation is Pareto optimal.

...le **10.E.1:** *The Welfare Effects of a Distortionary Tax.* Consider again the ... tax problem studied in Example 10.C.1. Suppose now that the welfare ... keeps a balanced budget and returns the tax revenue raised to consumers ... of lump-sum transfers. What impact does this tax-and-transfer scheme have ...fare?[20]

... answer this question, it is convenient to let $(x_1^*(t), \ldots, x_I^*(t), q_1^*(t), \ldots, q_J^*(t))$ ... denote the equilibrium consumption, production, and price levels of good ... the tax rate is t. Note that $\phi_i'(x_i^*(t)) = p^*(t) + t$ for all i and that ... $= p^*(t)$ for all j. Thus, letting $x^*(t) = \sum_i x_i^*(t)$ and $S^*(t) = S(x^*(t))$, we can ... to express the change in aggregate Marshallian surplus resulting from

... To see this, check first that $S''(x) \leq 0$ at all x. Hence, $S(\cdot)$ is a concave function and therefore ... maximizes aggregate surplus if and only if $S'(x^*) = 0$. Then verify that $S'(x) = P(x) - C'(x)$ ... > 0.

... This problem is closely related to that studied in Example 3.I.1 (we could equally well ... the analysis here by asking, as we did there, about the welfare cost of the distortionary ... to the use of a lump-sum tax that raises the same revenue; the measure of deadweight loss ...ges would be the same as that developed here). The discussion that follows amounts to ...on, in the quasilinear context, of the analysis of Example 3.I.1 to situations with many ...rs and the presence of firms. For an approach that uses the theory of a normative ...ative consumer presented in Section 4.D, see the small-type discussion at the end of this

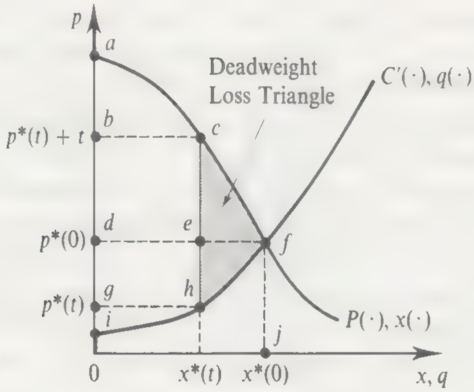


Figure 1̶0̶.̶E̶.̶3̶
The de... welfare ... distort...

the introduction of the tax as

$$S^*(t) - S^*(0) = \int_{x^*(0)}^{x^*(t)} [P(s) - C'(s)]\, ds. \tag{10.E.6}$$

Expression (10.E.6) is negative because $x^*(t) < x^*(0)$ (recall the analysis of Example 10.C.1) and $P(x) \geq C'(x)$ for all $x \leq x^*(0)$, with strict inequality for $x < x^*(0)$. Hence, social welfare is optimized by setting $t = 0$. The loss in welfare from $t > 0$ is known as the *deadweight loss of distortionary taxation* and is equal to the area of the shaded region in Figure 10.E.3, called the *deadweight loss triangle*.

Notice that since $S^{*\prime}(t) = [P(x^*(t)) - C'(x^*(t))]x^{*\prime}(t)$, we have $S^{*\prime}(0) = 0$. That is, starting from a position without any tax, the first-order welfare effect of an infinitesimal tax is zero. Only as the tax rate increases above zero does the marginal effect become strictly negative. This is as it should be: if we start at an (interior) welfare maximum, then a small displacement from the optimum cannot have a first-order effect on welfare.

It is sometimes of interest to distinguish between the various components of aggregate Marshallian surplus that accrue directly to consumers, firms, and the tax authority.[21] The *aggregate consumer surplus* when consumers' effective price is $\hat{p}$ and therefore aggregate consumption is $x(\hat{p})$ is defined as the gross consumer benefits from consumption of good ℓ minus the consumers' total expenditure on this good (the latter is the cost to consumers in terms of forgone consumption of the numeraire):

$$CS(\hat{p}) = \sum_{i=1}^{I} \phi_i(x_i(\hat{p})) - \hat{p}x(\hat{p}).$$

Using again the fact that consumption is distributed optimally, we have

$$\begin{aligned} CS(\hat{p}) &= \int_0^{x(\hat{p})} P(s)\, ds - \hat{p}x(\hat{p}) \\ &= \int_0^{x(\hat{p})} [P(s) - \hat{p}]\, ds. \end{aligned} \tag{10.E.7}$$

21. For example, if the set of active consumers of good ℓ is distinct from the set of owners of the firms producing the good, then this distinction tells us something about the distributional effects of the tax in the absence of transfers between owners and consumers.

...ly, the integral in (10.E.7) is equal to[22]

$$CS(\hat{p}) = \int_{\hat{p}}^{\infty} x(s)\,ds. \tag{10.E.8}$$

... because consumers face an effective price of $p^*(t) + t$ when the tax is t, the ...ge in consumer surplus from imposition of the tax is

$$CS(p^*(t) + t) - CS(p^*(0)) = -\int_{p^*(0)}^{p^*(t)+t} x(s)\,ds. \tag{10.E.9}$$

...gure 10.E.3, the reduction in consumer surplus is depicted by area $(dbcf)$.

...e aggregate profit, or *aggregate producer surplus*, when firms face effective ...e $\hat{p}$ is

$$\Pi(\hat{p}) = \hat{p}q(\hat{p}) - \sum_{j=1}^{J} c_j(q_j(\hat{p})).$$

..., using the optimality of the allocation of production across firms, we have[23]

$$\Pi(\hat{p}) = \Pi_0 + \int_0^{q(\hat{p})} [\hat{p} - C'(s)]\,ds \tag{10.E.10}$$

$$= \Pi_0 + \int_0^{\hat{p}} q(s)\,ds, \tag{10.E.11}$$

... Π_0 is a constant of integration equal to profits when $q_j = 0$ for all j [$\Pi_0 = 0$ ... $= 0$ for all j]. Since producers pay no tax, they face price $p^*(t)$ when the tax ... is t. The change in producer surplus is therefore

$$\Pi(p^*(t)) - \Pi(p^*(0)) = -\int_{p^*(t)}^{p^*(0)} q(s)\,ds. \tag{10.E.12}$$

...eduction in producer surplus is depicted by area $(gdfh)$ in Figure 10.E.3.

...ally, the *tax revenue* is $tx^*(t)$; it is depicted in Figure 10.E.3 by area $(gbch)$.

...e total deadweight welfare loss from the tax is then equal to the sum of the ...tions in consumer and producer surplus less the tax revenue. ■

... welfare measure developed here is closely related to our discussion of normative ...tative consumers in Section 4.D. We showed there that if a central authority is ...uting wealth to maximize a social welfare function given prices p, leading to a wealth ...tion rule $(w_1(p, w), \ldots, w_I(p, w))$, then there is a normative representative consumer ...direct utility function $v(p, w)$ whose demand $x(p, w)$ is exactly equal to aggregate ... [i.e., $x(p, w) = \sum_i x_i(p, w_i(p, w))$] and whose utility can be used as a measure of social .... Recalling our discussion in Section 3.I, this means that we can measure the change ...are resulting from a price-wealth change by adding the representative consumer's

22. This can be seen geometrically. For example, when $\hat{p} = p^*(0)$, the integrals in both (10.E.7) ... (10.E.8) are equal to area (daf) in Figure 10.E.3. Formally, the equivalence follows from a ... of variables and integration by parts (see Exercise 10.E.2).

23. When $\hat{p} = p^*(0)$, the integrals in both (10.E.10) and (10.E.11) are equal to area (idf) in ... 10.E.3. The equivalence of these two integrals again follows formally by a change of variables ... integration by parts.

compensating or equivalent variation for the price change to the change in the representative consumer's wealth (see Exercise 3.I.12). But in the quasilinear case, the representative consumer's compensating and equivalent variations are the same and can be calculated by direct integration of the representative consumer's Walrasian demand function, that is, by integration of the aggregate demand function. Hence, in Example 10.E.1, the representative consumer's compensating variation for the price change is exactly equal to the change in aggregate consumer surplus, expression (10.E.9). The change in the representative consumer's wealth, on the other hand, is equal to the change in aggregate profits plus the tax revenue rebated to consumers. Thus, the total welfare change arising from the introduction of the tax-and-transfer scheme, as measured using the normative representative consumer, is exactly equal to the deadweight loss calculated in Example 10.E.1.[24]

Another way to justify the use of aggregate surplus as a welfare measure in the quasilinear model is as a measure of *potential Pareto improvement*. Consider the tax example. We could say that a change in the tax represents a *potential* Pareto improvement if there is a set of lump-sum transfers of the numeraire that would make all consumers better off than they were before the tax change. In the present quasilinear context, this is true if and only if aggregate surplus increases with the change in the tax. This approach is sometimes referred to as the *compensation principle* because it asks whether, in principle, it is possible given the change for the winners to compensate the losers so that all are better off than before. (See also the discussion in Example 4.D.2 and especially Section 22.C.)

We conclude this section with a warning: When the numeraire represents many goods, the welfare analysis we have performed is justified only if the prices of goods other than good ℓ are undistorted in the sense that they equal these goods' true marginal utilities and production costs. Hence, these other markets must be competitive, and all market participants must face the same price. If this condition does not hold, then the costs of production faced by producers of good ℓ do not reflect the true social costs incurred from their use of these goods as inputs. Exercise 10.G.3 provides an illustration of this problem.

10.F Free-Entry and Long-Run Competitive Equilibria

Up to this point, we have taken the set of firms and their technological capabilities as fixed. In this section, we consider the case in which an infinite number of firms can potentially be formed, each with access to the most efficient production technology. Moreover, firms may enter or exit the market in response to profit opportunities. This scenario, known as a situation of *free entry*, is often a reasonable approximation when we think of long-run outcomes in a market. In the discussion that follows, we introduce and study a notion of *long-run competitive equilibrium* and then discuss how this concept can be used to analyze long-run and short-run comparative statics effects.

To begin, suppose that each of an infinite number of potential firms has access to a technology for producing good ℓ with cost function $c(q)$, where q is the *individual* firm's output of good ℓ. We assume that $c(0) = 0$; that is, a firm can earn zero profits by simply deciding to be inactive and setting $q = 0$. In the terminology of Section

24. This deadweight loss measure corresponds also to the measure developed for the one-consumer case in Example 3.I.1, where we implicitly limited ourselves to the case in which the taxed good has a constant unit cost.

are no sunk costs in the long run. The aggregate demand function is $x(\cdot)$, demand function $P(\cdot)$.

-run competitive equilibrium, we would like to determine not only the output levels for the firms but also the number of firms that are active in Given our assumption of identical firms, we focus on equilibria in which firms produce the same output level, so that a long-run competitive can be described by a triple (p, q, J) formed by a price p, an output per an integer number of active firms J (hence the total industry output is The central assumption determining the number of active firms is one of exit: A firm will enter the market if it can earn positive profits at the price and will exit if it can make only negative profits at any positive level given this price. If all firms, active and potential, take prices as their own actions, this implies that active firms must earn exactly zero long-run competitive equilibrium; otherwise, we would have either no to be active in the market (if profits were negative) or an infinite number the market (if profits were positive). This leads us to the formulation Definition 10.F.1.

F.1: Given an aggregate demand function $x(p)$ and a cost function $c(q)$ potentially active firm having $c(0) = 0$, a triple (p^*, q^*, J^*) is a *long-run equilibrium* if

$$q^* \text{ solves } \operatorname*{Max}_{q \geq 0}\; p^*q - c(q) \qquad \text{(Profit maximization)}$$

$$x(p^*) = J^*q^* \qquad \text{(Demand = supply)}$$

$$p^*q^* - c(q^*) = 0 \qquad \text{(Free Entry Condition).}$$

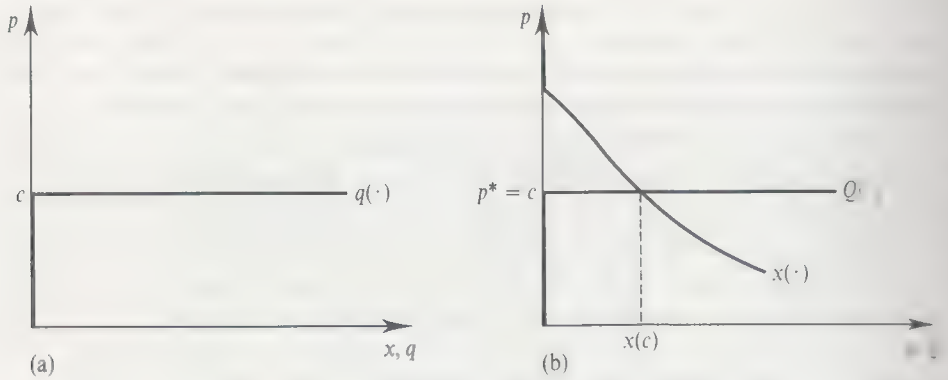
-run equilibrium price can be thought of as equating demand with where the long-run supply takes into account firms' entry and exit particular, if $q(\cdot)$ is the supply correspondence of an individual firm with and $\pi(\cdot)$ is its profit function, we can define a *long-run aggregate correspondence* by[26]

$$= \begin{cases} \infty & \text{if } \pi(p) > 0, \\ \{Q \geq 0\colon Q = Jq \text{ for some integer } J \geq 0 \text{ and } q \in q(p)\} & \text{if } \pi(p) = 0. \end{cases}$$

every firm wants to supply an amount strictly bounded away from aggregate supply is infinite. If $\pi(p) = 0$ and $Q = Jq$ for some $q \in q(p)$, have J firms each supply q and have the rest remain inactive [since a profit-maximizing choice for the inactive firms as well]. With this

assumption that all active firms produce the same output level is without loss of whenever $c(\cdot)$ is strictly convex on the set $(0, \infty]$. A firm's supply correspondence can at most one positive output level at any given price p.

the basic properties of production sets presented in Section 5.B, the long-run correspondence is the supply correspondence of the production set Y^+, where Y is the set associated with the individual firm [i.e., with $c(\cdot)$], and Y^+ is its "additive closure" set that contains Y and is additive: $Y^+ + Y^+ \subset Y^+$; see Exercise 5.B.4).



notion of a long-run supply correspondence, p^* is a long-run competitive equilibrium price if and only if $x(p^*) \in Q(p^*)$.[27]

We now investigate this long-run competitive equilibrium notion. Consider first the case in which the cost function $c(\cdot)$ exhibits constant returns to scale, so that $c(q) = cq$ for some $c > 0$, and assume that $x(c) > 0$. In this case, condition (i) of Definition 10.F.1 tells us that in any long-run competitive equilbrium we have $p^* \leq c$ (otherwise, there is no profit-maximizing production). However, at any such price aggregate consumption is strictly positive since $x(c) > 0$, so condition (ii) requires that $q^* > 0$. By condition (iii), we must have $(p^* - c)q^* = 0$. Hence, we conclude that $p^* = c$ and aggregate consumption is $x(c)$. Note, however, that J^* and q^* are *indeterminate*: any J^* and q^* such that $J^*q^* = x(c)$ satisfies conditions (i) and (ii).

Figure 10.F.1 depicts this long-run equilibrium. The supply correspondence of an individual firm $q(\cdot)$ is illustrated in Figure 10.F.1(a); Figure 10.F.1(b) shows the long-run equilibrium price and aggregate output as the intersection of the graph of the aggregate demand function $x(\cdot)$ with the graph of the long-run aggregate supply correspondence

$$Q(p) = \begin{cases} \infty & \text{if } p > c \\ [0, \infty) & \text{if } p = c \\ 0 & \text{if } p < c. \end{cases}$$

We move next to the case in which $c(\cdot)$ is increasing and strictly convex (i.e., the production technology of an individual firm displays strictly decreasing returns to scale). We assume also that $x(c'(0)) > 0$. With this type of cost function, *no long-run competitive equilibrium can exist*. To see why this is so, note that if $p > c'(0)$, then $\pi(p) > 0$ and therefore the long-run supply is infinite. On the other hand, if $p \leq c'(0)$, then the long-run supply is zero while $x(p) > 0$. The problem is illustrated in Figure 10.F.2, where the graph of the demand function $x(\cdot)$ has no intersection with the

27. In particular, if (p^*, q^*, J^*) is a long-run equilibrium, then condition (i) of Definition 10.F.1 implies that $q^* \in q(p^*)$ and condition (iii) implies that $\pi(p^*) = 0$. Hence, by condition (ii), $x(p^*) \in Q(p^*)$. In the other direction, if $x(p^*) \in Q(p^*)$, then $\pi(p^*) = 0$ and there exists $q^* \in q(p^*)$ and J^* with $x(p^*) = J^*q^*$. Therefore, the three conditions of Definition 10.F.1 are satisfied.

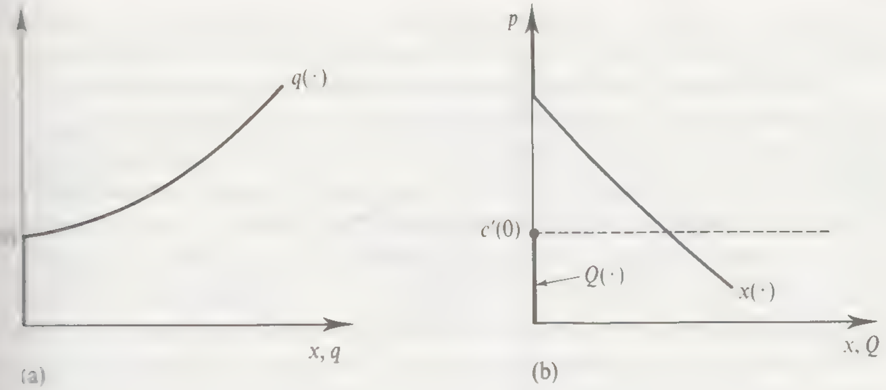


Figure 10.F.2
Nonexistence of long-run competitive equilibrium with strictly convex costs. (a) A firm's supply correspondence. (b) No intersection of long-run supply and demand.

of the long-run aggregate supply correspondence

$$Q(p) = \begin{cases} \infty & \text{if } p > c'(0) \\ 0 & \text{if } p \leq c'(0). \end{cases}$$

---

difficulty can be understood in a related way. As discussed in Exercise 5.B.4, the aggregate production set in the situation just described is convex but not closed. be seen in Figure 10.F.3, where the industry marginal cost function with J firms,

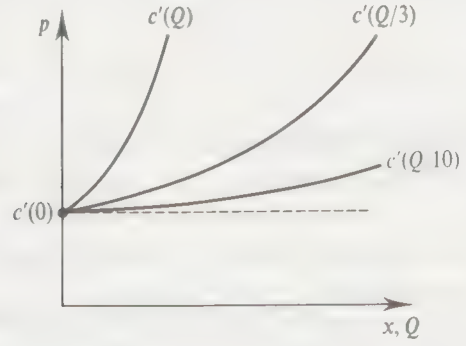


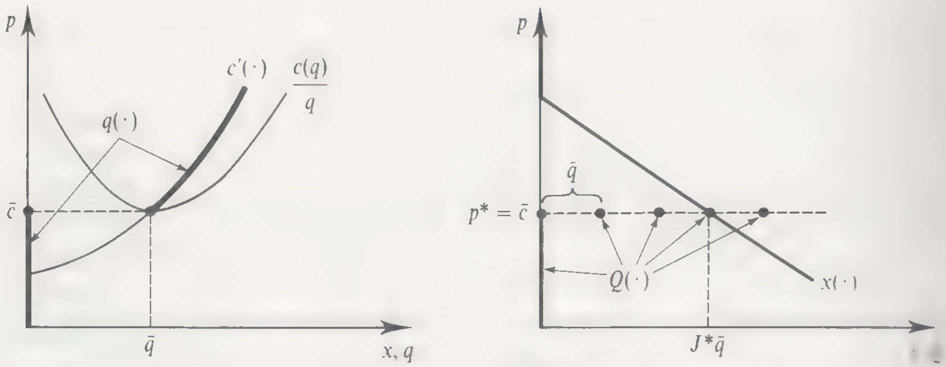
Figure 10.F.3
The limiting behavior of industry marginal cost as $J \to \infty$ with strictly convex costs.

is shown for various values of J (in particular, for $J = 1$, $J = 3$, and $J = 10$). Note ncreases, this marginal cost function approaches *but never reaches* the marginal cost corresponding to a constant marginal cost of $c'(0)$.

---

ps not surprisingly, to generate the existence of an equilibrium with a ate number of firms, the long-run cost function must exhibit a strictly fficient scale; that is, *there must exist a strictly positive output level $\bar{q}$ at which average costs of production are minimized* (see Section 5.D for a further sion of the efficient scale concept).

se, in particular, that $c(\cdot)$ has a unique efficient scale $\bar{q} > 0$, and let the d level of average cost be $\bar{c} = c(\bar{q})\ \bar{q}$. Assume, moreover, that $x(\bar{c}) > 0$. If at n equilibrium (p^*, q^*, J^*) we had $p^* > \bar{c}$, then $p^*\bar{q} > \bar{c}\bar{q}$, and so we would $) > 0$. Thus, at any long-run equilibrium we must have $p^* \leq \bar{c}$. In contrast, $< \bar{c}$, then $x(p^*) > 0$; but since $p^*q - c(q) = p^*q - (c(q)/q)q \leq (p^* - \bar{c})q < 0$



(a) (b)

for all $q > 0$, a firm would earn strictly negative profits at any positive level of output. So $p^* < \bar{c}$ also cannot be a long-run equilibrium price. Thus, at any long-run equilibrium we must have $p^* = \bar{c}$. Moreover, if $p^* = \bar{c}$, then each active firm's supply must be $q^* = \bar{q}$ (this is the only strictly positive output level at which the firm earns nonnegative profits), and the equilibrium number of active firms is therefore $J^* = x(\bar{c})/\bar{q}$.[28] In conclusion, the number of active firms is a well-determined quantity at long-run equilibrium. Figure 10.F.4 depicts such an equilibrium. The long-run aggregate supply correspondence is

$$Q(p) = \begin{cases} \infty & \text{if } p > \bar{c} \\ \{Q \geq 0: Q = J\bar{q} \text{ for some integer } J \geq 0\} & \text{if } p = \bar{c} \\ 0 & \text{if } p < \bar{c}. \end{cases}$$

Observe that the equilibrium price and aggregate output are exactly the same as if the firms had a constant returns to scale technology with unit cost $\bar{c}$.

Several points should be noted about the equilibrium depicted in Figure 10.F.4. First, if the efficient scale of operation is large relative to the size of market demand, it could well turn out that the equilibrium number of active firms is small. In these cases, we may reasonably question the appropriateness of the price-taking assumption (e.g., what if $J^* = 1$?). Indeed, we are then likely to be in the realm of the situations with market power studied in Chapter 12.

Second, we have conveniently shown the demand at price $\bar{c}$, $x(\bar{c})$, to be an integer multiple of $\bar{q}$. Were this not so, no long-run equilibrium would exist because the graphs of the demand function and the long-run supply correspondence would

28. Note that when $c(\cdot)$ is differentiable, condition (i) of Definition 10.F.1 implies that $c'(q^*) = p^*$, while condition (iii) implies $p^* = c(q^*)/q^*$. Thus, a *necessary* condition for an equilibrium is that $c'(q^*) = c(q^*)/q^*$. This is the condition for q^* to be a critical point of average costs [differentiate $c(q)/q$ and see Exercise 5.D.1]. In the case where average cost $c(q)/q$ is U-shaped (i.e., with no critical point other than the global minimum, as shown in Figure 10.F.4), this implies that $q^* = \bar{q}$, and so $p^* = \bar{c}$ and $J^* = x(\bar{c})/\bar{q}$. Note, however, that the argument in the text does not require this assumption about the shape of average costs.

[illegible] The nonexistence of competitive equilibrium can occur here for the [illegible] that we have already alluded to in small type in Section 10.C: The [illegible] production technologies we are considering exhibit nonconvexities.

[illegible] plausible, however, that when the efficient scale of a firm is small relative [illegible] the market, this "integer problem" should not be too much of a concern. [illegible] we study oligopolistic markets in Chapter 12, we shall see that when [illegible] scales are small in this sense, the oligopolistic equilibrium price is [illegible] equilibrium price we would derive if we simply ignored the integer [illegible] the number of firms J^*. Intuitively, when the efficient scale is small, [illegible] many firms in the industry and the equilibrium, although not strictly [illegible] will involve a price close to $\bar{c}$. Thus, if the efficient scale is small relative [illegible] the market [as measured by $x(\bar{c})$], then ignoring the integer problem [illegible] firms as price takers gives approximately the correct answer.

[illegible] an equilibrium exists, as in Figure 10.F.4, the equilibrium outcome [illegible] Marshallian aggregate surplus and therefore is Pareto optimal. To see [illegible] Figure 10.F.4 that aggregate surplus at the considered equilibrium is [illegible]

$$\operatorname*{Max}_{x \geq 0} \int_0^x P(s)\, ds - \bar{c}x,$$

[illegible] value of aggregate surplus when firms' cost functions are $\bar{c}q$. But [illegible] q for all q, this must be the largest attainable value of aggregate [illegible] given the actual cost function $c(\cdot)$; that is,

$$\operatorname*{Max}_{x \geq 0} \int_0^x P(s)\, ds - \bar{c}x \geq \int_0^{\hat{x}} P(s)\, ds - Jc(\hat{x}/J),$$

[illegible] J. This fact provides an example of a point we raised at the end of [illegible] (and will substantiate with considerable generality in Chapter 16): The [illegible] theorem continues to be valid even in the absence of convexity of [illegible] production sets.

[illegible] and Long-Run Comparative Statics

[illegible] firms may enter and exit the market in response to profit opportunities [illegible] these changes may take time. For example, factories may need to be [illegible] workforce reduced, and machinery sold when a firm exits an industry. [illegible] pay a firm to continue operating until a suitable buyer for its plant and [illegible] be found. When examining the comparative statics effects of a shock [illegible] is therefore important to distinguish between long-run and short-run [illegible]

[illegible] for example, that we are at a long-run equilibrium with J^* active firms

[illegible] intermediate case between constant returns (where any scale is efficient) and the case of [illegible] scale occurs when there is a range $[\underline{q}, \bar{q}]$ of efficient scales (the average cost curve [illegible] bottom). In this case, the integer problem is mitigated. For a long-run competitive [illegible] to exist, we now only need there to be some $q \in [\underline{q}, \bar{q}]$ such that $x(\bar{c})/q$ is an integer. [illegible] as the interval $[\underline{q}, \bar{q}]$ grows larger, not only are the chances of a long-run equilibrium [illegible] greater, but so are the chances of indeterminacy of the equilibrium number of firms (i.e., of [illegible] equilibria involving differing numbers of firms).

each producing q^* units of output and that there is some shock to demand (similar points can be made for supply shocks). In the short run, it may be impossible for any new firms to organize and enter the industry, and so we will continue to have J^* firms for at least some period of time. Moreover, these J^* firms may face a short-run cost function $c_s(\cdot)$ that differs from the long-run cost function $c(\cdot)$ because various input levels may be fixed in the short run. For example, firms may have the long-run cost function

$$c(q) = \begin{cases} K + \psi(q) & \text{if } q > 0 \\ 0 & \text{if } q = 0, \end{cases} \tag{10.F.1}$$

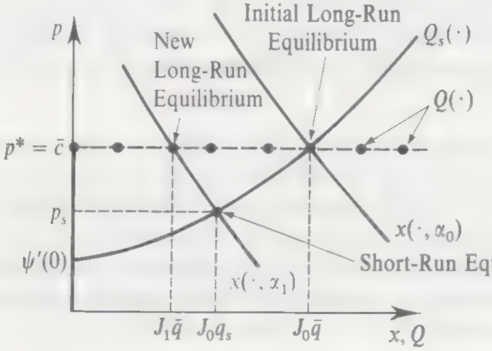
where $\psi(0) = 0$, $\psi'(q) > 0$, and $\psi''(q) > 0$. But in the short run, it may be impossible for an active firm to recover its fixed costs if it exits and sets $q = 0$. Hence, in the short run the firm has the cost function

$$c_s(q) = K + \psi(q) \qquad \text{for all } q \geq 0. \tag{10.F.2}$$

Another possibility is that $c(q)$ might be the cost function of some multiple-input production process, and in the short run an active firm may be unable to vary its level of some inputs. (See the discussion in Section 5.B on this point and also Exercises 10.F.5 and 10.F.6 for illustrations.)

Whenever the distinction between short run and long run is significant, the *short-run comparative statics effects* of a demand shock may best be determined by solving for the competitive equilibrium given J^* firms, each with cost function $c_s(\cdot)$, and the new demand function. This is just the equilibrium notion studied in Section 10.C, where we take firms' cost functions to be $c_s(\cdot)$. The *long-run comparative statics effects* can then be determined by solving for the long-run (i.e., free entry) equilibrium given the new demand function and long-run cost function $c(\cdot)$.

Example 10.F.1: *Short-Run and Long-Run Comparative Statics with Lumpy Fixed Costs that Are Sunk in the Short Run.* Suppose that the long-run cost function $c(\cdot)$ is given by (10.F.1) but that in the short run the fixed cost K is sunk so that $c_s(\cdot)$ is given by (10.F.2). The aggregate demand function is initially $x(\cdot, \alpha_0)$, and the industry is at a long-run equilibrium with J_0 firms, each producing $\bar{q}$ units of output [the efficient scale for cost function $c(\cdot)$], and a price of $p^* = \bar{c} = c(\bar{q})/\bar{q}$. This equilibrium position is depicted in Figure 10.F.5.



suppose that we have a shift to the demand function $x(\cdot, \alpha_1)$ shown in Figure ... The short-run equilibrium is determined by the intersection of the graph of ...mand function with the graph of the industry supply correspondence of the ... each of which has short-run cost function $c_s(\cdot)$. The short-run aggregate ...rrespondence is depicted as $Q_s(\cdot)$ in the figure. Thus, in the short run, the ... demand causes price to fall to p_s and output per firm to fall to q_s. Firms' ...s also fall; since $p_s < \bar{c}$, active firms lose money in the short run.

...e long run, however, firms exit in response to the decrease in demand, with ...ber of firms falling to $J_1 < J_0$, each producing output $\bar{q}$. Price returns to ... aggregate consumption is $x(\bar{c}, \alpha_1)$, and all active firms once again earn zero ...s. This new long-run equilibrium is also shown in Figure 10.F.5. ■

...vision of dynamic adjustment into two periods, although useful as a first approxima-...dmittedly crude. It may often be reasonable to think that there are several distinct ...stages corresponding to different levels of adjustment costs associated with different ...in the very short run, production may be completely fixed; in the medium run, some ...y be adjusted while others may not be; perhaps entry and exit take place only in ...ong run." Moreover, the methodology that we have discussed treats the two periods ... from each other. This approach ignores, for example, the possibility of intertemporal ...n by consumers when tomorow's price is expected to differ from today's (inter-... substitution might be particularly important for very short-run periods when the ...many production decisions are fixed can make prices very sensitive to demand shocks). ...weaknesses are not flaws in the competitive model per se, but rather only in the ... extreme methodological simplification adopted here. A fully satisfactory treatment ...ssues requires an explicitly dynamic model that places expectations at center stage. ...er 20 we study dynamic models of competitive markets in greater depth. Nevertheless, ...e dichotomization into long-run and short-run periods of adjustment is often a useful ... point for analysis.

...ncluding Remarks on Partial Equilibrium Analysis

...ciple, the analysis of Pareto optimal outcomes and competitive equilibria ...s the simultaneous consideration of the entire economy (a task we undertake ... IV). Partial equilibrium analysis can be thought of as facilitating matters on ...unts. On the positive side, it allows us to determine the equilibrium outcome ... particular market under study in isolation from all other markets. On the ...ve side, it allows us to use Marshallian aggregate surplus as a welfare measure ... many cases of interest, has a very convenient representation in terms of the ...ying vertically between the aggregate demand and supply curves.

...e model considered in Sections 10.C to 10.F, the validity of both of these ...ations rested, implicitly, on two premises: first, that the prices of all ...ities other than the one under consideration remain fixed; second, that there ... wealth effects in the market under study. We devote this section to a few ...al interpretative comments regarding these assumptions. (See also Section ... an example illustrating the limits of partial equilibrium analysis.)

The assumption that the prices of goods other than the good under consideration (say, good ℓ) remain fixed is essential for limiting our positive and normative analysis to a single market. In Section 10.B, we justified this assumption in terms of the market for good ℓ being small and having a diffuse influence over the remaining markets. However, this is not its only possible justification. For example, the nonsubstitution theorem (see Appendix A of Chapter 5) implies that the prices of all other goods will remain fixed if the numeraire is the only primary (i.e., nonproduced) factor, all produced goods other than ℓ are produced under conditions of constant returns using the numeraire and produced commodities other than ℓ as inputs, and there is no joint production.[30]

Even when we cannot assume that all other prices are fixed, however, a generalization of our single-market partial equilibrium analysis is sometimes possible. Often we are interested not in a single market but in a group of commodities that are strongly interrelated either in consumers' tastes (tea and coffee are the classic examples) or in firms' technologies. In this case, studying one market at a time while keeping other prices fixed is no longer a useful approach because what matters is the *simultaneous* determination of *all* prices in the group. However, if the prices of goods outside the group may be regarded as unaffected by changes within the markets for this group of commodities, and if there are no wealth effects for commodities in the group, then we can extend much of the analysis presented in Sections 10.C to 10.F.

To this effect, suppose that the group is composed of M goods, and let $x_i \in \mathbb{R}^M_+$ and $q_j \in \mathbb{R}^M$ be vectors of consumptions and productions for these M goods. Each consumer has a utility function of the form

$$u_i(m_i, x_i) = m_i + \phi_i(x_i),$$

where m_i is the consumption of the numeraire commodity (i.e., the total expenditure on commodities outside the group). Firms' cost functions are $c_j(q_j)$. With this specification, many of the basic results of the previous sections go through unmodified (often it is just a matter of reinterpreting x_i and q_j as vectors). In particular, the results discussed in Section 10.C on the uniqueness of equilibrium and its independence from initial endowments still hold (see Exercise 10.G.1), as do the welfare theorems of Section 10.D. However, our ability to conduct welfare analysis using the areas lying vertically between demand and supply curves becomes much more limited. The cross-effects among markets with changing and interrelated prices cannot be

30. A simple example of this result arises when all produced goods other than ℓ are produced directly from the numeraire with constant returns to scale. In this case, the equilibrium price of each of these goods is equal to the amount of the numeraire that must be used as an input in its production per unit of output produced. More generally, prices for produced goods other than ℓ will remain fixed under the conditions of the nonsubstitution theorem because all efficient production vectors can be generated using a single set of techniques. In any equilibrium, the price of each produced good other than ℓ must be equal to the amount of the numeraire embodied in a unit of the good in the efficient production technique, either directly through the use of the numeraire as an input or indirectly through the use as inputs of produced goods other than ℓ that are in turn produced using the numeraire (or using other produced goods that are themselves produced using the numeraire, and so on).

(Exercises 10.G.3 to 10.G.5 ask you to consider some issues related to this

...ssumption of no wealth effects for good ℓ, on the other hand, is critical for ...ty of the style of welfare analysis that we have carried out in this chapter. ... it, as we shall see in Part IV, Pareto optimality cannot be determined ...dently from the particular distribution of welfare sought, and we already ...m Section 3.I that area measures calculated from Walrasian demand ... are not generally correct measures of compensating or equivalent variations ...h the Hicksian demand functions should be used). However, the assumption ...ealth effects is much less critical for positive analysis (determination of ...m, comparative statics effects, and so on). Even with wealth effects, the ...-and-supply apparatus can still be quite helpful for the positive part of the ... The behavior of firms, for example, is not changed in any way. Consumers, ...ther hand, have a demand function that, with prices of the other goods ...ed, now depends only on the price for good ℓ and wealth. If wealth is ...ed from initial endowments and shareholdings, then we can view wealth as ...nction of the price of good ℓ (recall that other prices are fixed), and so we ... express demand as a function of this good's price alone. Formally, the ...educes to that presented in Section 10.C: The equilibrium in market ℓ can ...ntified as an intersection point of demand and supply curves.[32]

...ase in which the single-market analysis for good ℓ is still fully justified is when utility ...nctions have the form

$$u_i(m_i, x_i) = m_i + \phi_{\ell i}(x_{\ell i}) + \phi_{-\ell,i}(x_{-\ell,i}),$$

$$c_j(q_j) = c_{\ell j}(q_{\ell j}) + c_{-\ell,j}(q_{-\ell,j}),$$

... $x_{-\ell,i}$ and $q_{-\ell,j}$ are consumption and production vectors for goods in the group other than ... additive separability in good ℓ, the markets for goods in the group other than ℓ do ...nce the equilibrium price in market ℓ. Good ℓ is effectively independent of the group, and ...eat it in isolation, as we have done in the previous sections. (In point of fact, we do not ...ssume that the remaining markets in the group keep their prices fixed. What happens ... is simply irrelevant for equilibrium and welfare analysis in the market for good ℓ.) See ... G.2.

...he presence of wealth effects can lead, however, to some interesting new phenomena on ...'s side. One is the *backward-bending* demand curve, where demand for a good is ... in its price over some range. This can happen if consumers have endowments of good ℓ, ...hen an increase in its price increases consumers' wealth and could lead to a net increase ...mands for good ℓ, even if it is a normal good.

...RENCES

... (1920). *Principles of Economics.* New York: Macmillan.

...). *The Theory of Price*, 4th ed. New York: Macmillan.

...). Small income effects: A Marshallian theory of consumer surplus and downward sloping ...d. *Review of Economic Studies* **54**: 87–103.

EXERCISES

10.B.1[B] The concept defined in Definition 10.B.2 is sometimes known as *strong Pareto efficiency*. An outcome is *weakly Pareto efficient* if there is no alternative feasible allocation that makes *all* individuals *strictly* better off.

(a) Argue that if an outcome is strongly Pareto efficient, then it is weakly Pareto efficient as well.

(b) Show that if all consumers' preferences are continuous and strongly monotone, then these two notions of Pareto efficiency are equivalent for any *interior* outcome (i.e., an outcome in which each consumer's consumption lies in the interior of his consumption set). Assume for simplicity that $X_i = \mathbb{R}^L_+$ for all i.

(c) Construct an example where the two notions are not equivalent. Why is the strong monotonicity assumption important in **(b)**? What about interiority?

10.B.2[A] Show that if allocation $(x_1^*, \ldots, x_I^*, y_1^*, \ldots, y_J^*)$ and price vector $p^* \gg 0$ constitute a competitive equilibrium, then allocation $(x_1^*, \ldots, x_I^*, y_1^*, \ldots, y_J^*)$ and price vector αp^* also constitute a competitive equilibrium for any scalar $\alpha > 0$.

10.C.1[B] Suppose that consumer i's preferences can be represented by the utility function $u_i(x_{1i}, \ldots, x_{Li}) = \sum_\ell \log(x_{\ell i})$ (these are Cobb–Douglas preferences).

(a) Derive his demand for good ℓ. What is the wealth effect?

(b) Now consider a sequence of situations in which we proportionately increase both the number of goods and the consumer's wealth. What happens to the wealth effect in the limit?

10.C.2[B] Consider the two-good quasilinear model presented in Section 10.C with one consumer and one firm (so that $I = 1$ and $J = 1$). The initial endowment of the numeraire is $\omega_m > 0$, and the initial endowment of good ℓ is 0. Let the consumer's quasilinear utility function be $\phi(x) + m$, where $\phi(x) = \alpha + \beta \ln x$ for some $(\alpha, \beta) \gg 0$. Also, let the firm's cost function be $c(q) = \sigma q$ for some scalar $\sigma > 0$. Assume that the consumer receives all the profits of the firm. Both the firm and the consumer act as price takers. Normalize the price of good m to equal 1, and denote the price of good ℓ by p.

(a) Derive the consumer's and the firm's first-order conditions.

(b) Derive the competitive equilibrium price and output of good ℓ. How do these vary with α, β, and σ?

10.C.3[B] Consider a central authority who operates J firms with differentiable convex cost functions $c_j(q_j)$ for producing good ℓ from the numeraire. Define $C(q)$ to be the central authority's minimized cost level for producing aggregate quantity q; that is

$$C(q) = \underset{(q_1, \ldots, q_J) \geq 0}{\text{Min}} \quad \sum_{j=1}^{J} c_j(q_j)$$
$$\text{s.t.} \quad \sum_{j=1}^{J} q_j \geq q.$$

(a) Derive the first-order conditions for this cost-minimization problem.

(b) Show that at the cost-minimizing production allocation $(q_1^*, \ldots, q_J^*)$, $C'(q) = c_j'(q_j^*)$ for all j with $q_j^* > 0$ (i.e., the central authority's marginal cost at aggregate output level q equals each firm's marginal cost level at the optimal production allocation for producing q).

(c) Show that if firms all maximize profit facing output price $p = C'(q)$ (with the price of the numeraire equal to 1), then the consequent output choices result in an aggregate output of q. Conclude that $C'(\cdot)$ is the inverse of the industry supply function $q(\cdot)$.

Consider a central authority who has x units of good ℓ to allocate among I consumers, of whom has a quasilinear utility function of the form $\phi_i(x_i) + m_i$, with $\phi_i(\cdot)$ a ...able, increasing, and strictly concave function. The central authority allocates good ...aximize the sum of consumers' utilities $\sum_i u_i$.

(a) Set up the central authority's problem and derive its first-order condition.

... $\gamma(x)$ be the value function of the central authority's problem, and let $P(x) = \gamma'(x)$ ...vative. Show that if $(x_1^*, \ldots, x_I^*)$ is the optimal allocation of good ℓ given available ... x, then $P(x) = \phi_i'(x_i^*)$ for all i with $x_i^* > 0$.

...gue that if all consumers maximize utility facing a price for good ℓ of $P(x)$ (with the ...e numeraire equal to 1), then the aggregate demand for good ℓ is exactly x. Conclude ... $P(\cdot)$ is, in fact, the inverse of the aggregate demand function $x(\cdot)$.

Derive the differential change in the equilibrium price in response to a differential ... the tax in Example 10.C.1 by applying the implicit function theorem to the system ...ations (10.C.4) to (10.C.6).

A tax is to be levied on a commodity bought and sold in a competitive market. ...e forms of tax may be used: In one case, a *specific* tax is levied, where an amount ... per unit bought or sold (this is the case considered in the text); in the other case, an ... tax is levied, where the government collects a tax equal to τ times the amount the ...receives from the buyer. Assume that a partial equilibrium approach is valid.

... that, with a specific tax, the ultimate cost of the good to consumers and the ...urchased are independent of whether the consumers or the producers pay the tax.

... that this is not generally true with an ad valorem tax. In this case, which collection ... to a higher cost to consumers? Are there special cases in which the collection ...s irrelevant with an ad valorem tax?

An ad valorem tax of τ (see Exercise 10.C.6 for a definition) is to be levied on ... a competitive market with aggregate demand curve $x(p) = Ap^{\varepsilon}$, where $A > 0$ ... $\varepsilon < 0$, and aggregate supply curve $q(p) = \alpha p^{\gamma}$, where $\alpha > 0$ and $\gamma > 0$. Calculate the ... change in consumer cost and producer receipts per unit sold for a small ...al") tax. Denote $\kappa = (1 + \tau)$. Assume that a partial equilibrium approach is valid. ... the elasticity of the equilibrium price with respect to κ. Argue that when $\gamma = 0$ ...ear the full effect of the tax while consumers' total costs of purchase are unaffected, ...when $\varepsilon = 0$ it is consumers who bear the full burden of the tax. What happens when ... these elasticities approaches ∞ in absolute value?

...ppose that there are J firms producing good ℓ, each with a differentiable cost ... $c(q, \alpha)$ that is strictly convex in q, where α is an exogenous parameter that affects ... be a technological parameter or an input price). Assume that $\partial c(q, \alpha)/\partial\alpha > 0$. ...able aggregate demand function for good ℓ is $x(p)$, with $x'(\cdot) \le 0$. Assume that ...equilibrium analysis is justified.

... be the *per firm* output and $p^*(\alpha)$ be the equilibrium price in the competitive ... given α.

Derive the marginal change in a firm's profits with respect to α.

... the weakest possible sufficient condition, stated in terms of marginal and average ... their derivatives, that guarantees that if α increases marginally, then firms' ... profits decline for any demand function $x(\cdot)$ having $x'(\cdot) \le 0$. Show that if this ... is not satisfied, then there are demand functions such that profits increase when α

(c) In the case where α is the price of factor input k, interpret the condition in **(b)** in terms of the conditional factor demand for input k.

10.C.9^{B} Suppose that in a partial equilibrium context there are J identical firms that produce good ℓ with cost function $c(w, q)$, where w is a vector of factor input prices. Show that an increase in the price of factor k, w_k, lowers the equilibrium price of good ℓ if and only if factor k is an *inferior* factor, that is, if at fixed input prices, the use of factor k is decreasing in a firm's output level.

10.C.10^{B} Consider a market with demand curve $x(p) = \alpha p^{\varepsilon}$ and with J firms, each of which has marginal cost function $c'(q) = \beta q^{\eta}$, where $(\alpha, \beta, \eta) \gg 0$ and $\varepsilon < 0$. Calculate the competitive equilibrium price and output levels. Examine the comparative statics change in these variables as a result of changes in α and β. How are these changes affected by ε and η?

10.C.11^{B} Assume that partial equilibrium analysis is valid. Suppose that firms 1 and 2 are producing a positive level of output in a competitive equilibrium. The cost function for firm j is given by $c(q, \alpha_j)$, where α_j is an exogenous technological parameter. If α_1 differs from α_2 marginally, what is the difference in the two firms' profits?

10.D.1^{B} Prove that under the assumptions that the $\phi_i(\cdot)$ functions are strictly concave and the cost functions $c_j(\cdot)$ are convex, the optimal individual consumption levels of good ℓ in problem (10.D.2) are uniquely defined. Conclude that the optimal aggregate production level of good ℓ is therefore also uniquely defined. Show that if the cost functions $c_j(\cdot)$ are strictly convex, then the optimal individual production levels of good ℓ in problem (10.D.2) are also uniquely defined.

10.D.2^{B} Determine the optimal consumption and production levels of good ℓ for the economy described in Exercise 10.C.2. Compare these with the equilibrium levels you identified in that exercise.

10.D.3^{B} In the context of the two-good quasilinear economy studied in Section 10.D, show that any allocation that is a solution to problem (10.D.6) is Pareto optimal and that any Pareto optimal allocation is a solution to problem (10.D.6) for *some* choice of utility levels $(\bar{u}_2, \ldots, \bar{u}_I)$.

10.D.4^{B} Derive the first-order conditions for problem (10.D.6) and compare them with conditions (10.D.3) to (10.D.5).

10.E.1^{C} Suppose that $J_d > 0$ of the firms that produce good ℓ are domestic firms, and $J_f > 0$ are foreign firms. All domestic firms have the same convex cost function for producing good ℓ, $c_d(q_j)$. All foreign firms have the same convex cost function $c_f(q_j)$. Assume that partial equilibrium analysis is valid.

The government of the domestic country is considering imposing a per-unit tariff of τ on imports of good ℓ. The government wants to maximize domestic welfare as measured by the *domestic* Marshallian surplus (i.e., the sum of domestic consumers' utilities less domestic firms' costs).

(a) Show that if $c_f(\cdot)$ is strictly convex, then imposition of a small tariff raises domestic welfare.

(b) Show that if $c_f(\cdot)$ exhibits constant returns to scale, then imposition of a small tariff lowers domestic welfare.

10.E.2^{B} Consumer surplus when consumers face effective price $\hat{p}$ can be written as

$$CS(\hat{p}) = \int_0^{x(\hat{p})} [P(s) - \hat{p}]\, ds.$$

... by means of a change of variables and integration by parts that this integral is equal ... $\int_{\bar{p}} x(s)\,ds$.

... 3[c] (*Ramsey tax problem*) Consider a fully separable quasilinear model with L goods in ... each consumer has preferences of the form $u_i(x_i) = x_{1i} + \sum_{\ell=2}^{L} \phi_{\ell i}(x_{\ell i})$ and each good ... L is produced with constant returns to scale from good 1, using c_ℓ units of good 1 per ... good ℓ produced. Assume that consumers initially hold endowments only of the ...aire, good 1. Hence, consumers are net sellers of good 1 to the firms and net purchasers ... goods $2, \ldots, L$.

... setting, consumer i's demand for each good $\ell \neq 1$ can be written in the form $x_{\ell i}(p_\ell)$, ... demand for good ℓ is independent of the consumer's wealth and all other prices, and ... can be measured by the sum of the Marshallian aggregate surpluses in the $L-1$ ... for nonnumeraire commodities (see Section 10.G and Exercise 10.G.2 for more ...).

...pose that the government must raise R units of good 1 through (specific) commodity ... Note, in particular, that such taxes involve taxing a *transaction* of a good, *not* an ...dual's consumption level of that good.

... denote the tax to be paid by a consumer in units of good 1 for each unit of good ... purchased, and let t_1 be the tax in units of good 1 to be paid by consumers for each ... good 1 sold to a firm. Normalize the price paid by firms for good 1 to equal 1. Under ...assumptions, each choice of $t = (t_1, \ldots, t_L)$ results in a consumer paying a total of $c_\ell + t_\ell$ ... of good $\ell \neq 1$ purchased and having to part with $(1 + t_1)$ units of good 1 for each ... of good 1 sold to a firm.

(a) Consider two possible tax vectors t and t'. Show that if t' is such that $(c_\ell + t'_\ell) =$ ... $+ t_\ell)$ and $(1 + t'_1) = (1/\alpha)(1 + t_1)$ for some scalar $\alpha > 0$, then the two sets of taxes raise ... revenue. Conclude from this fact that the government can restrict attention to tax ...s that leave one good untaxed.

... Let good 1 be the untaxed good (i.e., set $t_1 = 0$). Derive conditions describing the taxes ...ld be set on goods $2, \ldots, L$ if the government wishes to minimize the welfare loss ... from this taxation. Express this formula in terms of the elasticity of demand for each ...

... Under what circumstances should the tax rate on all goods be equal? In general, which ... should have higher tax rates? When would taxing only good 1 be optimal?

...1[a] Show that if $c(q)$ is strictly convex in q and $c(0) = 0$, then $\pi(p) > 0$ if and only if ...

...2[a] Consider a market with demand function $x(p) = A - Bp$ in which every potential firm ... function $c(q) = K + \alpha q + \beta q^2$, where $\alpha > 0$ and $\beta > 0$.

... Calculate the long-run competitive equilibrium price, output per firm, aggregate ... and number of firms. Ignore the integer constraint on the number of firms. How does ... of these vary with A?

... Now examine the short-run competitive equilibrium response to a change in A starting ... long-run equilibrium you identified in (a). How does the change in price depend on ... of A in the initial equilibrium? What happens as $A \to \infty$? What accounts for this ... of market size?

... (D. Pearce) Consider a partial equilibrium setting in which each (potential) firm has ... cost function $c(\cdot)$, where $c(q) = K + \phi(q)$ for $q > 0$ and $c(0) = 0$. Assume that ... and $\phi''(q) < 0$, and denote the firm's efficient scale by $\bar{q}$. Suppose that there is initially ... equilibrium with J^* firms. The government considers imposing two different types

of taxes: The first is an ad valorem tax of τ (see Exercise 10.C.6) on sales of the good. The second is a tax T that must be paid by any operating firm (where a firm is considered to be "operating" if it sells a positive amount). If the two taxes would raise an equal amount of revenue with the initial level of sales and number of firms, which will raise more after the industry adjusts to a new long-run equilibrium? (You should ignore the integer constraint on the number of firms.)

10.F.4$^{\mathrm{B}}$ (J. Panzar) Assume that partial equilibrium analysis is valid. The single-output, many-input technology for producing good ℓ has a differentiable cost function $c(w, q)$, where $w = (w_1, \ldots, w_K)$ is a vector of factor input prices and q is the firm's output of good ℓ. Given factor prices w, let $\bar{q}(w)$ denote the firm's efficient scale. Assume that $\bar{q}(w) > 0$ for all w. Also let $p_\ell^*(w)$ denote the long-run equilibrium price of good ℓ when factor prices are w. Show that the function $p_\ell^*(w)$ is nondecreasing, homogeneous of degree one, and concave. (You should ignore the integer constraint on the number of firms.)

10.F.5$^{\mathrm{C}}$ Suppose that there are J firms that can produce good ℓ from K factor inputs with differentiable cost function $c(w, q)$. Assume that this function is strictly convex in q. The differentiable aggregate demand function for good ℓ is $x(p, \alpha)$, where $\partial x(p, \alpha)/\partial p < 0$ and $\partial x(p, \alpha)/\partial \alpha > 0$ (α is an exogenous parameter affecting demand). However, although $c(w, q)$ is the cost function when all factors can be freely adjusted, factor k cannot be adjusted in the short run.

Suppose that we are initially at an equilibrium in which all inputs are optimally adjusted to the equilibrium level of output q^* and factor prices w so that, letting $z_k(w, q)$ denote a firm's conditional factor demand for input k when all inputs can be adjusted, $z_k^* = z_k(w, q^*)$.

(a) Show that a firm's equilibrium response to an increase in the price of good ℓ is larger in the long run than in the short run.

(b) Show that this implies that the long-run equilibrium response of p_ℓ to a marginal increase in α is smaller than the short-run response. Show that the reverse is true for the response of the equilibrium aggregate consumption of good ℓ (hold the number of firms equal to J in both the short run and long run).

10.F.6$^{\mathrm{B}}$ Suppose that the technology for producing a good uses capital (z_1) and labor (z_2) and takes the Cobb–Douglas form $f(z_1, z_2) = z_1^\alpha z_2^{1-\alpha}$, where $\alpha \in (0, 1)$. In the long run, both factors can be adjusted; but in the short run, the use of capital is fixed. The industry demand function takes the form $x(p) = a - bp$. The vector of input prices is (w_1, w_2). Find the long-run equilibrium price and aggregate quantity. Holding the number of firms and the level of capital fixed at their long-run equilibrium levels, what is the short-run industry supply function?

10.F.7$^{\mathrm{B}}$ Consider a case where in the short run active firms can increase their use of a factor but cannot decrease it. Show that the short-run cost curve will exhibit a kink (i.e., be nondifferentiable) at the current (long-run) equilibrium. Analyze the implications of this fact for the relative variability of short-run prices and quantities.

10.G.1$^{\mathrm{B}}$ Consider the case of an interrelated group of M commodities. Let consumer i's utility function take the form $u_i(x_{1i}, \ldots, x_{Mi}) = m_i + \phi_i(x_{1i}, \ldots, x_{Mi})$. Assume that $\phi_i(\cdot)$ is differentiable and strictly concave. Let firm j's cost function be the differentiable convex function $c_j(q_{1j}, \ldots, q_{Mj})$.

Normalize the price of the numeraire to be 1. Derive $(I + J + 1)M$ equations characterizing the $(I + J + 1)M$ equilibrium quantities $(x_{1i}^*, \ldots, x_{Mi}^*)$ for $i = 1, \ldots, I$, $(q_{1j}^*, \ldots, q_{Mj}^*)$ for $j = 1, \ldots, J$, and $(p_1^*, \ldots, p_M^*)$. [*Hint*: Derive consumers' and firms' first-order conditions and the $M - 1$ market-clearing conditions in parallel to our analysis of the single-market case.] Argue that the equilibrium prices and quantities of these M goods are independent of

consumers' wealths, that equilibrium individual consumptions and aggregate production levels are unique, and that if the $c_j(\cdot)$ functions are strictly convex, then equilibrium individual production levels are also unique.

10.G.2[B] Consider the case in which the functions $\phi_i(\cdot)$ and $c_j(\cdot)$ in Exercise 10.G.1 are separable in good ℓ (one of the goods in the group): $\phi_i(\cdot) = \phi_{\ell i}(x_{\ell i}) + \phi_{-\ell,i}(x_{-\ell,i})$ and $c_j(\cdot) = c_{\ell j}(q_{\ell j}) + c_{-\ell,j}(q_{-\ell,j})$. Argue that in this case, the equilibrium price, consumption, and production of good ℓ can be determined independently of other goods in the group. Also argue that under the same assumptions as in the single-market case studied in Section 10.E, changes in welfare caused by changes in the market for this good can be captured by the Marshallian aggregate surplus for this good, $\sum_i \phi_{\ell i}(x_{\ell i}) - \sum_j c_{\ell j}(q_{\ell j})$, which can be represented in terms of the areas lying vertically between the demand and supply curves for good ℓ. Note the implication of these results for the case in which we have separability of all goods: $\phi_i(\cdot) = \sum_\ell \phi_{\ell i}(x_{\ell i})$ and $c_j(\cdot) = \sum_\ell c_{\ell j}(q_{\ell j})$.

10.G.3[B] Consider a three-good economy ($\ell = 1, 2, 3$) in which every consumer has preferences that can be described by the utility function $u(x) = x_1 + \phi(x_2, x_3)$ and there is a single production process that produces goods 2 and 3 from good 1 having $c(q_2, q_3) = c_2 q_2 + c_3 q_3$. Suppose that we are considering a tax change in only a single market, say market 2.

(a) Show that if the price in market 3 is undistorted (i.e., if $t_3 = 0$), then the change in aggregate surplus caused by the tax change can be captured solely through the change in the area lying vertically between market 2's demand and supply curves holding the price of good 3 at its initial level.

(b) Show that if market 3 is initially distorted because $t_3 > 0$, then by using only the single-market measure in **(a)**, we would overstate the decrease in aggregate surplus if good 3 is a substitute for good 2 and would understate it if good 3 is a complement. Provide an intuitive explanation of this result. What is the correct measure of welfare change?

10.G.4[B] Consider a three-good economy ($\ell = 1, 2, 3$) in which every consumer has preferences that can be described by the utility function $u(x) = x_1 + \phi(x_2, x_3)$ and there is a single production process that produces goods 2 and 3 from good 1 having $c(q_2, q_3) = c_2 q_2 + c_3 q_3$. Derive an expression for the welfare loss from an increase in the tax rates on both goods.

10.G.5[B] Consider a three-good economy ($\ell = 1, 2, 3$) in which every consumer has preferences that can be described by the utility function $u(x) = x_1 + \phi(x_2, x_3)$ and there is a single production process that produces goods 2 and 3 from good 1 having $c(q_2, q_3) = c_2(q_2) + c_3(q_3)$, where $c_2(\cdot)$ and $c_3(\cdot)$ are strictly increasing and strictly convex.

(a) If goods 2 and 3 are substitutes, what effect does an increase in the tax on good 2 have on the price paid by consumers for good 3? What if they are complements?

(b) What is the bias from applying the formula for welfare loss you derived in part (b) of Exercise 10.G.3 using the price paid by consumers for good 3 prior to the tax change in both the case of substitutes and that of complements?

CHAPTER

11 Externalities and Public Goods

11.A Introduction

In Chapter 10, we saw a close connection between competitive, price-taking equilibria and Pareto optimality (or, Pareto efficiency).[1] The first welfare theorem tells us that competitive equilibria are necessarily Pareto optimal. From the second welfare theorem, we know that under suitable convexity hypotheses, any Pareto optimal allocation can be achieved as a competitive allocation after an appropriate lump-sum redistribution of wealth. Under the assumptions of these theorems, the possibilities for welfare-enhancing intervention in the marketplace are strictly limited to the carrying out of wealth transfers for the purposes of achieving distributional aims.

With this chapter, we begin our study of *market failures*: situations in which some of the assumptions of the welfare theorems do *not* hold and in which, as a consequence, market equilibria cannot be relied on to yield Pareto optimal outcomes. In this chapter, we study two types of market failure, known as *externalities* and *public goods*.

In Chapter 10, we assumed that the preferences of a consumer were defined solely over the set of goods that she might herself decide to consume. Similarly, the production of a firm depended only on its own input choices. In reality, however, a consumer or firm may in some circumstances be directly affected by the actions of other agents in the economy; that is, there may be *external effects* from the activities of other consumers or firms. For example, the consumption by consumer i's neighbor of loud music at three in the morning may prevent her from sleeping. Likewise, a fishery's catch may be impaired by the discharges of an upstream chemical plant. Incorporating these concerns into our preference and technology formalism is, in principle, a simple matter: We need only define an agent's preferences or production set over both her own actions and those of the agent creating the external effect. But the effect on market equilibrium is significant: In general, when external effects are present, competitive equilibria are not Pareto optimal.

Public goods, as the name suggests, are commodities that have an inherently "public" character, in that consumption of a unit of the good by one agent does not preclude its consumption by another. Examples abound: Roadways, national defense,

1. See also Chapter 16.

...rol projects, and knowledge all share this characteristic. The private ... of public goods generates a special type of externality: if one individual ... unit of a public good, all individuals benefit. As a result, private provision ... goods is typically Pareto inefficient.

...gin our investigation of externalities and public goods in Section 11.B by ... the simplest possible externality: one that involves only two agents in ...my, where one of the agents engages in an activity that directly affects the ... this setting, we illustrate the inefficiency of competitive equilibria when an ... is present. We then go on to consider three traditional solutions to this ... quotas, taxes, and the fostering of decentralized bargaining over the extent ...xternality. The last of these possibilities also suggests a connection between ...ce of externalities and the nonexistence of certain commodity markets, a ... we explore in some detail.

...ction 11.C, we study public goods. We first derive a condition that ...es the optimal level of a public good and we then illustrate the inefficiency ...rom private provision. This Pareto inefficiency can be seen as arising from ...ity among the consumers of the good, which in this context is known as ...er *problem*. We also discuss possible solutions to this free-rider problem. ...tity-based intervention (here, direct governmental provision) and price- ...rvention (taxes and subsidies) can, in principle, correct it. In contrast, ...ed bargaining and competitive market-based solutions are unlikely to be ... the context of public goods.

...tion 11.D, we return to the analysis of externalities. We study cases in which ...s both produce and are affected by the externality. Multilateral externalities ...ssified according to whether the externality is *depletable* (or *private* or ...or *nondepletable* (or *public* or *nonrivalrous*). We argue that market ... are likely to work well in the former set of cases but poorly in the latter, ... externality possesses the characteristics of a public good (or bad). Indeed, ... well explain why most externalities that are regarded as serious social ... e.g., water pollution, acid rain, congestion) take the form of nondepletable ...ral externalities.

... Section 11.E, we examine another problem that may arise in these settings: ...als may have privately held information about the effects of externalities ... well-being. We see there that this type of informational asymmetry may ... both private and government efforts to achieve optimal outcomes.

...pendix A, we study the connection between externalities and the presence ...gical nonconvexities, and we examine the implications of these nonconvexi- ... our analysis.

... literature on externalities and public goods is voluminous. Useful introduc- ...and further references to these subjects may be found in Baumol and Oates ... and Laffont (1988).

...mple Bilateral Externality

...ingly, perhaps, a fully satisfying definition of an externality has proved ...hat elusive. Nevertheless, informal Definition 11.B.1 provides a serviceable ... of departure.

Definition 11.B.1: An *externality* is present whenever the well-being of a consumer or the production possibilities of a firm are directly affected by the actions of another agent in the economy.

Simple as Definition 11.B.1 sounds, it contains a subtle point that has been a source of some confusion. When we say "directly," we mean to exclude any effects that are mediated by prices. That is, an externality is present if, say, a fishery's productivity is affected by the emissions from a nearby oil refinery, but not simply because the fishery's profitability is affected by the price of oil (which, in turn, is to some degree affected by the oil refinery's output of oil). The latter type of effect [referred to as a *pecuniary externality* by Viner (1931)] is present in any competitive market but, as we saw in Chapter 10, creates no inefficiency. Indeed, with price-taking behavior, the market is precisely the mechanism that guarantees a Pareto optimal outcome. This suggests that the presence of an externality is not merely a technological phenomenon but also a function of the set of markets in existence. We return to this point later in the section.

In the remainder of this section, we explore the implications of external effects for competitive equilibria and public policy in the context of a very simple two-agent, partial equilibrium model. We consider two consumers, indexed by $i = 1, 2$, who constitute a small part of the overall economy. In line with this interpretation, we suppose that the actions of these consumers do not affect the prices $p \in \mathbb{R}^L$ of the L traded goods in the economy. At these prices, consumer i's wealth is w_i.

In contrast with the standard competitive model, however, we assume that each consumer has preferences not only over her consumption of the L traded goods $(x_{1i}, \ldots, x_{Li})$ but also over some action $h \in \mathbb{R}_+$ taken by consumer 1. Thus, consumer i's (differentiable) utility function takes the form $u_i(x_{1i}, \ldots, x_{Li}, h)$, and we assume that $\partial u_2(x_{12}, \ldots, x_{L2}, h)/\partial h \neq 0$. Because consumer 1's choice of h affects consumer 2's well-being, it generates an externality. For example, the two consumers may live next door to each other, and h may be a measure of how loudly consumer 1 plays music. Or the consumers may live on a river, with consumer 1 further upstream. In this case, h could represent the amount of pollution put into the river by consumer 1; more pollution lowers consumer 2's enjoyment of the river. We should hasten to add that external effects need not be detrimental to those affected by them. Action h could, for example, be consumer 1's beautification of her property, which her neighbor, consumer 2, also gets to enjoy.[2]

In what follows, it will be convenient to define for each consumer i a derived utility function over the level of h, assuming optimal commodity purchases by consumer i at prices $p \in \mathbb{R}^L$ and wealth w_i:

$$v_i(p, w_i, h) = \underset{x_i \geq 0}{\text{Max}} \quad u_i(x_i, h)$$
$$\text{s.t. } p \cdot x_i \leq w_i.$$

For expositional purposes, we shall also assume that the consumers' utility functions

2. An externality favorable to the recipient is usually called a *positive externality*, and conversely for a *negative externality*.

take a quasilinear form with respect to a numeraire commodity (we comment below, in small type, on the simplifications afforded by this assumption). In this case, we can write the derived utility function $v_i(\cdot)$ as $v_i(p, w_i, h) = \phi_i(p, h) + w_i$.[3] Since prices of the L traded goods are assumed to be unaffected by any of the changes we are considering, we shall suppress the price vector p and simply write $\phi_i(h)$. We assume that $\phi_i(\cdot)$ is twice differentiable with $\phi_i''(\cdot) < 0$. Be warned, however, that the concavity assumption is less innocent than it looks: see Appendix A for further discussion of this point.

Although we shall speak in terms of this consumer interpretation, everything we do here applies equally well to the case in which the two agents are firms (or, for that matter, one firm and one consumer). For example, we could consider a firm j that has a derived profit function $\pi_j(p, h)$ over h given prices p. Suppressing the price vector p, the firm's profit can be written as $\pi_j(h)$, which plays the same role as the function $\phi_i(h)$ in the analysis that follows.

Nonoptimality of the Competitive Outcome

Suppose that we are at a competitive equilibrium in which commodity prices are p. That is, at the equilibrium position, each of the two consumers maximizes her utility limited only by her wealth and the prices p of the traded goods. It must therefore be the case that consumer 1 chooses her level of $h \geq 0$ to maximize $\phi_1(h)$. Thus, the equilibrium level of h, h^*, satisfies the necessary and sufficient first-order condition

$$\phi_1'(h^*) \leq 0, \quad \text{with equality if } h^* > 0. \tag{11.B.1}$$

For an interior solution, we therefore have $\phi_1'(h^*) = 0$.

In contrast, in any Pareto optimal allocation, the optimal level of h, h°, must maximize the *joint surplus* of the two consumers, and so must solve[4]

$$\underset{h \geq 0}{\text{Max}} \quad \phi_1(h) + \phi_2(h).$$

This problem gives us the necessary and sufficient first-order condition for h° of

$$\phi_1'(h^\circ) \leq -\phi_2'(h^\circ), \text{ with equality if } h^\circ > 0. \tag{11.B.2}$$

Hence, for an interior solution to the Pareto optimality problem, $\phi_1'(h^\circ) = -\phi_2'(h^\circ)$.

When external effects are present, so that $\phi_2'(h) \neq 0$ at all h, the equilibrium level of h is not optimal unless $h^\circ = h^* = 0$. Consider, for example, the case in which we have interior solutions, that is, where $(h^*, h^\circ) \gg 0$. If $\phi_2'(\cdot) < 0$, so that h generates

3. Indeed, suppose that $u_i(x_i, h) = g_i(x_{-1i}, h) + x_{1i}$, where x_{-1i} is consumer i's consumption of traded goods other than good 1. Then, the consumer's Walrasian demand function for these $L - 1$ traded goods, $x_{-1i}(\cdot)$, is independent of her wealth, and $v_i(p, w_i, h) = g_i(x_{-1i}(p, h), h) - p \cdot x_{-1i}(p, h) + w_i$. Thus, denoting $\phi_i(p, h) = g_i(x_{-1i}(p, h), h) - p \cdot x_{-1i}(p, h)$, we have obtained the desired form.

4. Recall the reasoning of Sections 10.D and 10.E, or note that at any Pareto optimal allocation in which h° is the level of h and w_i is consumer i's wealth level for $i = 1, 2$, it must be impossible to change h and reallocate wealth so as to make one consumer better off without making the other worse off. Thus, $(h^\circ, 0)$ must solve $\text{Max}_{h,T}\ \phi_1(h) + w_1 - T$ subject to $\phi_2(h) + w_2 + T \geq \bar{u}_2$, for some $\bar{u}_2$. Because the constraint holds with equality in any solution to this problem, substituting from the constraint for T in the objective function shows that h° must maximize the joint surplus of the two consumers $\phi_1(h) + \phi_2(h)$.

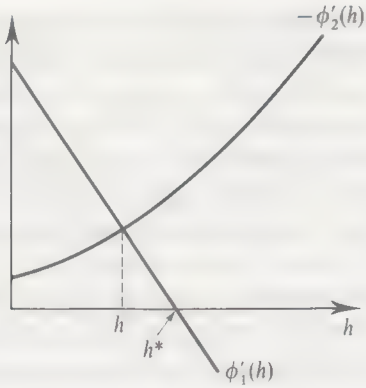


Figure 11.B.1
The equilibr
and optimal
of a negativ
externality.

a negative externality, then we have $\phi_1'(h^\circ) = -\phi_2'(h^\circ) > 0$; because $\phi_1'(\cdot)$ is decreasing and $\phi_1'(h^*) = 0$, this implies that $h^* > h^\circ$. In contrast, when $\phi_2'(\cdot) > 0$, h represents a positive externality, and $\phi_1'(h^\circ) = -\phi_2'(h^\circ) < 0$ implies that $h^* < h^\circ$.

Figure 11.B.1 depicts the solution for a case in which h constitutes a negative external effect, so that $\phi_2'(h) < 0$ at all h. In the figure, we graph $\phi_1'(\cdot)$ and $-\phi_2'(\cdot)$. The competitive equilibrium level of the externality h^* occurs at the point where the graph of $\phi_1'(\cdot)$ crosses the horizontal axis. In contrast, the optimal externality level h° corresponds to the point of intersection between the graphs of the two functions.

Note that optimality does not usually entail the complete elimination of a negative externality. Rather, the externality's level is adjusted to the point where the marginal benefit to consumer 1 of an additional unit of the externality-generating activity, $\phi_1'(h^\circ)$, equals its marginal cost to consumer 2, $-\phi_2'(h^\circ)$.

In the current example, quasilinear utilities lead the optimal level of the externality to be independent of the consumers' wealth levels. In the absence of quasilinearity, however, wealth effects for the consumption of the externality make its optimal level depend on the consumers' wealths. See Exercise 11.B.2 for an illustration. Note, however, that when the agents under consideration are firms, wealth effects are always absent.

Traditional Solutions to the Externality Problem

Having identified the inefficiency of the competitive market outcome in the presence of an externality, we now consider three possible solutions to the problem. We first look at government-implemented quotas and taxes, and then analyze the possibility that an efficient outcome can be achieved in a much less intrusive manner by simply fostering bargaining between the consumers over the extent of the externality.

Quotas and taxes

To fix ideas, suppose that h generates a negative external effect, so that $h^\circ < h^*$. The most direct sort of government intervention to achieve efficiency is the direct control of the externality-generating activity itself. The government can simply mandate that h be no larger than h°, its optimal level. With this constraint, consumer 1 will indeed fix the level of the externality at h°.

A second option is for the government to attempt to restore optimality by imposing a tax on the externality-generating activity. This solution is known as

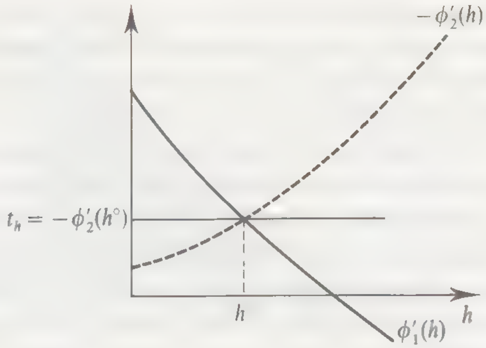


Figure 11.B.2
The optimality-restoring Pigouvian tax.

Pigouvian taxation, after Pigou (1932). To this effect, suppose that consumer 1 is made to pay a tax of t_h per unit of h. It is then not difficult to see that a tax of

$$t_h = -\phi_2'(h^\circ) > 0$$

will implement the optimal level of the externality. Indeed, consumer 1 will then choose the level of h that solves

$$\underset{h \geq 0}{\text{Max}} \quad \phi_1(h) - t_h h, \tag{11.B.3}$$

which has the necessary and sufficient first-order condition

$$\phi_1'(h) \leq t_h, \quad \text{with equality if } h > 0. \tag{11.B.4}$$

Given $t_h = -\phi_2'(h^\circ)$, $h = h^\circ$ satisfies condition (11.B.4) [recall that h° is defined by this condition: $\phi_1'(h^\circ) \leq -\phi_2'(h^\circ)$, with equality if $h^\circ > 0$]. Moreover, given $\phi_1''(\cdot) < 0$, h° must be the unique solution to problem (11.B.3). Figure 11.B.2 illustrates this solution for a case in which $h^\circ > 0$.

Note that the optimality-restoring tax is exactly equal to the *marginal externality* at the optimal solution.[5] That is, it is exactly equal to the amount that consumer 2 would be willing to pay to reduce h slightly from its optimal level h°. When faced with this tax, consumer 1 is effectively led to carry out an individual cost–benefit computation that *internalizes* the externality that she imposes on consumer 2.

The principles for the case of a positive externality are exactly the same, only now when we set $t_h = -\phi_2'(h^\circ) < 0$, t_h takes the form of a per-unit *subsidy* (i.e., consumer 1 *receives* a payment for each unit of the externality she generates).

Several additional points are worth noting about this Pigouvian solution. First, we can actually achieve optimality either by taxing the externality or by subsidizing its reduction. Consider, for example, the case of a negative externality. Suppose the government pays a subsidy of $s_h = -\phi_2'(h^\circ) > 0$ for every unit that consumer 1's choice of h is below h^*, its level in the competitive equilibrium. If so, then consumer 1 will maximize $\phi_1(h) + s_h(h^* - h) = \phi_1(h) - t_h h + t_h h^*$. But this is equivalent to a tax of t_h per unit on h combined with a lump-sum payment of $t_h h^*$. Hence, a subsidy for the reduction of the externality combined with a lump-sum transfer can exactly replicate the outcome of the tax.

Second, a point implicit in the derivation above is that, in general, it is essential

5. In the case where $h^\circ = 0$, any tax greater than $-\phi_2'(0)$ also implements the optimal outcome.

to tax the externality-producing activity directly. For instance, suppose that, in the example of consumer 1 playing loud music, we tax purchases of music equipment instead of taxing the playing of loud music itself. In general, this will not restore optimality. Consumer 1 will be led to lower her consumption of music equipment (perhaps she will purchase only a CD player, rather than a CD player and a tape player) but may nevertheless play whatever equipment she does purchase too loudly. A common example of this sort arises when a firm pollutes in the process of producing output. A tax on its output leads the firm to reduce its output level but may not have any effect (or, more generally, may have too little effect) on its pollution emissions. Taxing output achieves optimality only in the special case in which emissions bear a fixed monotonic relationship to the level of output. In this special case, emissions can be measured by the level of output, and a tax on output is essentially equivalent to a tax on emissions. (See Exercise 11.B.5 for an illustration.)

Third, note that the tax/subsidy and the quota approaches are equally effective in achieving an optimal outcome. However, the government must have a great deal of information about the benefits and costs of the externality for the two consumers to set the optimal levels of either the quota or the tax. In Section 11.E we will see that when the government does not possess this information the two approaches typically are not equivalent.

Fostering bargaining over externalities: enforceable property rights

Another approach to the externality problem aims at a less intrusive form of intervention, merely seeking to insure that conditions are met for the parties to themselves reach an optimal agreement on the level of the externality.

Suppose that we establish enforceable property rights with regard to the externality-generating activity. Say, for example, that we assign the right to an "externality-free" environment to consumer 2. In this case, consumer 1 is unable to engage in the externality-producing activity without consumer 2's permission. For simplicity, imagine that the bargaining between the parties takes a form in which consumer 2 makes consumer 1 a take-it-or-leave-it offer, demanding a payment of T in return for permission to generate externality level h.[6] Consumer 1 will agree to this demand if and only if she will be at least as well off as she would be by rejecting it, that is, if and only if $\phi_1(h) - T \geq \phi_1(0)$. Hence, consumer 2 will choose her offer (h, T) to solve

$$\begin{aligned} \operatorname*{Max}_{h \geq 0,\, T} \quad & \phi_2(h) + T \\ \text{s.t. } & \phi_1(h) - T \geq \phi_1(0). \end{aligned}$$

Since the constraint is binding in any solution to this problem, $T = \phi_1(h) - \phi_1(0)$. Therefore, consumer 2's optimal offer involves the level of h that solves

$$\operatorname*{Max}_{h \geq 0} \quad \phi_2(h) + \phi_1(h) - \phi_1(0). \tag{11.B.5}$$

But this is precisely h°, the socially optimal level.

Note, moreover, that the precise allocation of these rights between the two

6. Either of the bargaining processes discussed in Appendix A of Chapter 9 would yield the same conclusions.

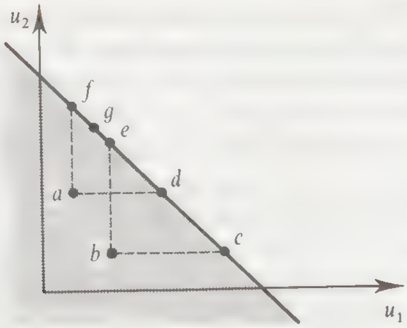


Figure 11.B.3 The final distribution of utilities under different property rights institutions and different bargaining procedures.

...mers is inessential to the achievement of optimality. Suppose, for example, that ...mer 1 instead has the right to generate as much of the externality as she wants. ... case, in the absence of any agreement, consumer 1 will generate externality ... h^*. Now consumer 2 will need to offer a $T < 0$ (i.e., to pay consumer 1) to have ... h^*. In particular, consumer 1 will agree to externality level h if and only if ... $- T \geq \phi_1(h^*)$. As a consequence, consumer 2 will offer to set h at the level that ... $\text{Max}_{h \geq 0} (\phi_2(h) + \phi_1(h) - \phi_1(h^*))$. Once again, the optimal externality level ...lts. The allocation of rights affects only the final wealth of the two consumers ...ering the payment made by consumer 1 to consumer 2. In the first case, ...mer 1 pays $\phi_1(h^\circ) - \phi_1(0) > 0$ to be allowed to set $h^\circ > 0$; in the second, she ...ys" $\phi_1(h^\circ) - \phi_1(h^*) < 0$ in return for setting $h^\circ < h^*$.

... have here an instance of what is known as the *Coase theorem* [for Coase ...]: If trade of the externality can occur, then bargaining will lead to an efficient ...come no matter how property rights are allocated.

... this is illustrated in Figure 11.B.3, in which we represent the utility possibility ... the two consumers. Every point in the boundary of this set corresponds to ...cation with externality level h°. The points a and b correspond to the utility ... arising, respectively, from externality levels 0 and h^* in the absence of any ...s. They constitute the initial situation after the assignment of property rights ...nsumers 2 and 1, respectively) but before bargaining. In the particular ...ning procedure we have adopted (which gives the power to make a take-it-or-...t offer to consumer 2), the utility levels after bargaining are points f and ...ectively. If the bargaining power (i.e., the power to make the take-it-or-leave-it ...) had been instead in the hands of consumer 1, the post-bargaining utility ... would have been points d and c, respectively. Other bargaining procedures ... the ones studied in Appendix A of Chapter 9) may yield other points in the ...ents $[f, d]$ and $[e, c]$, respectively.

... that the existence of both well-defined and enforceable property rights is ...al for this type of bargaining to occur. If property rights are not well defined, ... unclear whether consumer 1 must gain consumer 2's permission to generate ...ternality. If property rights cannot be enforced (perhaps the level of h is not ... measured), then consumer 1 has no need to purchase the right to engage in ...ernality-generating activity from consumer 2. For this reason, proponents of ... of approach focus on the absence of these legal institutions as a central ...diment to optimality.

This solution to the externality problem has a significant advantage over the tax and quota schemes in terms of the level of knowledge required of the government. The consumers must know each other's preferences, but the government need not. We should emphasize, however, that for bargaining over the externality to lead to efficiency, it is important that the consumers know this information. In Section 11.E, we will see that when the agents are to some extent ignorant of each others' preferences, bargaining need *not* lead to an efficient outcome.

Two further points regarding these three types of solutions to the externality problem are worthy of note. First, in the case in which the two agents are firms, one form that an efficient bargain might take is the sale of one of the firms to the other. The resulting merged firm would then fully internalize the externality in the process of maximizing its profits.[7]

Second, note that all three approaches require that the externality-generating activity be measureable. This is not a trivial requirement; in many cases, such measurement may be either technologically infeasible or very costly (consider the cost of measuring air pollution or noise). A proper computation of costs and benefits should take these costs into account. If measurement is very costly, then it may be optimal to simply allow the externality to persist.

Externalities and Missing Markets

The observation that bargaining can generate an optimal outcome suggests a connection between externalities and missing markets. After all, a market system can be viewed as a particular type of trading procedure.

Suppose that property rights are well defined and enforceable and that a competitive market for the right to engage in the externality-generating activity exists. For simplicity, we assume that consumer 2 has the right to an externality-free environment. Let p_h denote the price of the right to engage in one unit of the activity. In choosing how many of these rights to purchase, say h_1, consumer 1 will solve

$$\underset{h_1 \geq 0}{\text{Max}}\ \phi_1(h_1) - p_h h_1,$$

which has the first-order condition

$$\phi_1'(h_1) \leq p_h, \quad \text{with equality if } h_1 > 0. \tag{11.B.6}$$

In deciding how many rights to sell, h_2, consumer 2 will solve

$$\underset{h_2 \geq 0}{\text{Max}}\ \ \phi_2(h_2) + p_h h_2,$$

which has the first-order condition

$$\phi_2'(h_2) \leq -p_h, \quad \text{with equality if } h_2 > 0. \tag{11.B.7}$$

---

7. Note, however, that this conclusion presumes that the owner of a firm has full control over all its functions. In more complicated (but realistic) settings in which this is not true, say because owners must hire managers whose actions cannot be perfectly controlled, the results of a merger and of an agreement over the level of the externality need not be the same. Chapters 14 and 23 provide an introduction to the topic of incentive design. See Holmstrom and Tirole (1989) for a discussion of these issues in the theory of the firm.

In a competitive equilibrium, the market for these rights must clear; that is, we must have $h_1 = h_2$. Hence, (11.B.6) and (11.B.7) imply that the level of rights traded in this competitive rights market, say h^{**}, satisfies

$$\phi_1'(h^{**}) \leq -\phi_2'(h^{**}), \quad \text{with equality if } h^{**} > 0.$$

Comparing this expression with (11.B.2), we see that h^{**} equals the optimal level h°. The equilibrium price of the externality is $p_h^* = \phi_1'(h^\circ) = -\phi_2'(h^\circ)$.

Consumer 1 and 2's equilibrium utilities are then $\phi_1(h^\circ) - p_h^* h^\circ$ and $\phi_2(h^\circ) + p_h^* h^\circ$, respectively. The market therefore works as a particular bargaining procedure for splitting the gains from trade; for example, point g in Figure 11.B.3 could represent the utilities in the competitive equilibrium.

We see that if a competitive market exists for the externality, then optimality results. Thus, externalities can be seen as being inherently tied to the absence of certain competitive markets, a point originally noted by Meade (1952) and substantially extended by Arrow (1969). Indeed, recall that our original definition of an externality, Definition 11.B.1, explicitly required that an action chosen by one agent must *directly* affect the well-being or production capabilities of another. Once a market exists for an externality, however, each consumer decides for herself how much of the externality to consume at the going prices.

Unfortunately, the idea of a competitive market for the externality in the present example is rather unrealistic; in a market with only one seller and one buyer, price taking would be unlikely.[8] However, most important externalities are produced and felt by *many* agents. Thus, we might hope that in these multilateral settings, price taking would be a more reasonable assumption and, as a result, that a competitive market for the externality would lead to an efficient outcome. In Section 11.D, where we study multilateral externalities, we see that the correctness of this conclusion depends on whether the externality is "private" or "public" in nature. Before coming to this, however, we first study the nature of public goods.

11.C Public Goods

In this section, we study commodities that, in contrast with those considered so far, have a feature of "publicness" to their consumption. These commodities are known as *public goods*.

Definition 11.C.1: A *public good* is a commodity for which use of a unit of the good by one agent does not preclude its use by other agents.

Put somewhat differently, public goods possess the feature that they are *nondepletable*: Consumption by one individual does not affect the supply available for other individuals. Knowledge provides a good illustration. The use of a piece of knowledge for one purpose does not preclude its use for others. In contrast, the commodities studied up to this point have been assumed to be of a *private*, or *depletable*, nature;

8. For that matter, the idea that the externality rights are all sold at the same price lacks motivation here, because there is no natural unit of measurement for the externality.

that is, for each additional unit consumed by individual i, there is one unit less available for individuals $j \neq i$.[9]

A distinction can also be made according to whether *exclusion* of an individual from the benefits of a public good is possible. Every private good is automatically excludable, but public goods may or may not be. The patent system, for example, is a mechanism for excluding individuals (although imperfectly) from the use of knowledge developed by others. On the other hand, it might be technologically impossible, or at the least very costly, to exclude some consumers from the benefits of national defense or of a project to improve air quality. For simplicity, our discussion here will focus primarily on the case in which exclusion is not possible.

Note that a public "good" need not necessarily be desirable; that is, we may have public *bads* (e.g., foul air). In this case, we should read the phrase "does not preclude" in Definition 11.C.1 to mean "does not decrease."

Conditions for Pareto Optimality

Consider a setting with I consumers and one public good, in addition to L traded goods of the usual, private, kind. We again adopt a partial equilibrium perspective by assuming that the quantity of the public good has no effect on the prices of the L traded goods and that each consumer's utility function is quasilinear with respect to the same numeraire, traded commodity. As in Section 11.B, we can therefore define, for each consumer i, a derived utility function over the level of the public good. Letting x denote the quantity of the public good, we denote consumer i's utility from the public good by $\phi_i(x)$. We assume that this function is twice differentiable, with $\phi_i''(x) < 0$ at all $x \geq 0$. Note that precisely because we are dealing with a public good, the argument x does not have an i subscript.

The cost of supplying q units of the public good is $c(q)$. We assume that $c(\cdot)$ is twice differentiable, with $c''(q) > 0$ at all $q \geq 0$.

To describe the case of a desirable public good whose production is costly, we take $\phi_i'(\cdot) > 0$ for all i and $c'(\cdot) > 0$. Except where otherwise noted, however, the analysis applies equally well to the case of a public bad whose reduction is costly, where $\phi_i'(\cdot) < 0$ for all i and $c'(\cdot) < 0$.

In this quasilinear model, any Pareto optimal allocation must maximize aggregate surplus (see Section 10.D) and therefore must involve a level of the public good that solves

$$\underset{q \geq 0}{\text{Max}} \quad \sum_{i=1}^{I} \phi_i(q) - c(q).$$

The necessary and sufficient first-order condition for the optimal quantity q° is then

$$\sum_{i=1}^{I} \phi_i'(q^\circ) \leq c'(q^\circ), \quad \text{with equality if } q^\circ > 0. \tag{11.C.1}$$

Condition (11.C.1) is the classic optimality condition for a public good first derived by Samuelson (1954; 1955). (Here it is specialized to the partial equilibrium setting;

9. Intermediate cases are also possible in which the consumption of the good by one individual affects to some degree its availability to others. A classic example is the presence of congestion effects. For this reason, goods for which there is no depletability whatsoever are sometimes referred to as *pure* public goods.

See Section 16.G for a more general treatment.) At an interior solution, we have $\sum_i \phi_i'(q^\circ) = c'(q^\circ)$, so that at the optimal level of the public good *the sum of consumers' marginal benefits from the public good is set equal to its marginal cost*. This condition should be contrasted with conditions (10.D.3) to (10.D.5) for a private good, where *each* consumer's marginal benefit from the good is equated to its marginal cost.

Inefficiency of Private Provision of Public Goods

Consider the circumstance in which the public good is provided by means of private purchases by consumers. We imagine that a market exists for the public good and that each consumer i chooses how much of the public good to buy, denoted by $x_i \geq 0$, taking as given its market price p. The total amount of the public good purchased by consumers is then $x = \sum_i x_i$. Formally, we treat the supply side as consisting of a single profit-maximizing firm with cost function $c(\cdot)$ that chooses its production level taking the market price as given. Note, however, that as the analysis of Section 5.E, we can actually think of the supply behavior of this firm as representing the industry supply of J price-taking firms whose aggregate cost function is $c(\cdot)$.

At a competitive equilibrium involving price p^*, each consumer i's purchase of the public good x_i^* must maximize her utility and so must solve

$$\underset{x_i \geq 0}{\text{Max}} \quad \phi_i(x_i + \sum_{k \neq i} x_k^*) - p^* x_i. \tag{11.C.2}$$

In determining her optimal purchases, consumer i takes as given the amount of the public good being purchased by each other consumer (as in the Nash equilibrium concept studied in Section 8.D). Consumer i's purchases x_i^* must therefore satisfy the necessary and sufficient first-order condition

$$\phi_i'(x_i^* + \sum_{k \neq i} x_k^*) \leq p^*, \quad \text{with equality if } x_i^* > 0.$$

Letting $x^* = \sum_i x_i^*$ denote the equilibrium level of the public good, for each consumer i we must therefore have

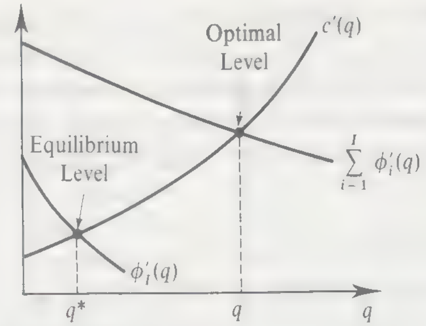
$$\phi_i'(x^*) \leq p^*, \quad \text{with equality if } x_i^* > 0. \tag{11.C.3}$$

The firm's supply q^*, on the other hand, must solve $\text{Max}_{q \geq 0}\,(p^* q - c(q))$ and therefore must satisfy the standard necessary and sufficient first-order condition

$$p^* \leq c'(q^*), \quad \text{with equality if } q^* > 0. \tag{11.C.4}$$

At a competitive equilibrium, $q^* = x^*$. Thus, letting $\delta_i = 1$ if $x_i^* > 0$ and $\delta_i = 0$ if $x_i^* = 0$, (11.C.3) and (11.C.4) tell us that $\sum_i \delta_i[\phi_i'(q^*) - c'(q^*)] = 0$. Recalling that $\phi_i'(\cdot) > 0$ and $c'(\cdot) > 0$, this implies that whenever $I > 1$ and $q^* > 0$ (so that $\delta_i = 1$ for some i) we have

$$\sum_{i=1}^{I} \phi_i'(q^*) > c'(q^*). \tag{11.C.5}$$



Comparing (11.C.5) with (11.C.1), we see that whenever $q^\circ > 0$ and $I > 1$, the level of the public good provided is too low; that is, $q^* < q^\circ$.[10]

The cause of this inefficiency can be understood in terms of our discussion of externalities in Section 11.B. Here each consumer's purchase of the public good provides a direct benefit not only to the consumer herself but also to every other consumer. Hence, private provision creates a situation in which externalities are present. The failure of each consumer to consider the benefits for others of her public good provision is often referred to as the *free-rider problem*: Each consumer has an incentive to enjoy the benefits of the public good provided by others while providing it insufficiently herself.

In fact, in the present model, the free-rider problem takes a very stark form. To see this most simply, suppose that we can order the consumers according to their marginal benefits, in the sense that $\phi_1'(x) < \cdots < \phi_I'(x)$ at all $x \geq 0$. Then condition (11.C.3) can hold with equality only for a *single* consumer and, moreover, this must be the consumer labeled I. Therefore, only the consumer who derives the largest (marginal) benefit from the public good will provide it; all others will set their purchases equal to zero in the equilibrium. The equilibrium level of the public good is then the level q^* that satisfies $\phi_I'(q^*) = c'(q^*)$. Figure 11.C.1 depicts both this equilibrium and the Pareto optimal level. Note that the curve representing $\sum_i \phi_i'(q)$ geometrically corresponds to a *vertical* summation of the individual curves representing $\phi_i(q)$ for $i = 1, \ldots, I$ (whereas in the case of a private good, the market demand curve is identified by adding the individual demand curves *horizontally*).

The inefficiency of private provision is often remedied by governmental intervention in the provision of public goods. Just as with externalities, this can happen not only through quantity-based intervention (such as direct governmental provision) but also through "price-based" intervention in the form of taxes or subsidies. For example, suppose that there are two consumers with benefit functions $\phi_1(x_1 + x_2)$ and $\phi_2(x_1 + x_2)$, where x_i is the amount of the public good purchased by consumer i, and that $q^\circ > 0$. By analogy with the analysis in Section 11.B, a subsidy to each consumer i per unit purchased of $s_i = \phi'_{-i}(q^\circ)$ [or, equivalently, a tax of $-\phi'_{-i}(q^\circ)$ per unit that consumer i's purchases of the public good fall below some specified

10. The conclusion follows immediately if $q^* = 0$. So suppose instead that $q^* > 0$. Then since $\sum_i \phi_i'(q^*) - c'(q^*) > 0$ and $\sum_i \phi_i'(\cdot) - c'(\cdot)$ is decreasing, any solution to (11.C.1) must have a larger value than q^*. Note that, in contrast, if we are dealing with a public bad, so that $\phi_i'(\cdot) < 0$ and $c'(\cdot) < 0$, then the inequalities reverse and $q^\circ < q^*$.

...el] faces each consumer with the marginal external effect of her actions and so ...erates an optimal level of public good provision by consumer i. Formally, if ... $\tilde{x}_2$) are the competitive equilibrium levels of the public good purchased by the ... consumers given these subsidies, and if $\tilde{p}$ is the equilibrium price, then consumer ... purchases of the public good, $\tilde{x}_i$, must solve $\text{Max}_{x_i \geq 0}\ \phi_i(x_i + \tilde{x}_j) + s_i x_i - \tilde{p} x_i$, and ... must satisfy the necessary and sufficient first-order condition

$$\phi_i'(\tilde{x}_1 + \tilde{x}_2) + s_i \leq \tilde{p}, \text{ with equality of } \tilde{x}_i > 0.$$

...stituting for s_i, and using both condition (11.C.4) and the market-clearing ...ition that $\tilde{x}_1 + \tilde{x}_2 = \tilde{q}$, we conclude that $\tilde{q}$ is the total amount of the public ... in the competitive equilibrium given these subsidies if and only if

$$\phi_i'(\tilde{q}) + \phi'_{-i}(q^\circ) \leq c'(\tilde{q}),$$

... equality for some i if $\tilde{q} > 0$. Recalling (11.C.1) we see that $\tilde{q} = q^\circ$. (Exercise 11.C.1 ... you to extend this argument to the case where $I > 2$; formally, we then have a ...ltilateral externality of the sort studied in Section 11.D.)

...te that both optimal direct public provision and this subsidy scheme require ... the government know the benefits derived by consumers from the public ... i.e., their willingness to pay in terms of private goods). In Section 11.E, we ... the case in which this is not so.

...dahl Equilibria

...gh private provision of the sort studied above results in an inefficient level of ... ...blic good, there is *in principle* a market institution that can achieve optimality. ...se that, for each consumer i, we have a market for the public good "as ...enced by consumer i." That is, we think of each consumer's consumption ... public good as a distinct commodity with its own market. We denote the price ... personalized good by p_i. Note that p_i may differ across consumers. Suppose ...at, given the equilibrium price p_i^{**}, each consumer i sees herself as deciding ...*the total amount of the public good she will consume*, x_i, so as to solve

$$\underset{x_i \geq 0}{\text{Max}} \quad \phi_i(x_i) - p_i^{**} x_i.$$

...ilibrium consumption level x_i^{**} must therefore satisfy the necessary and ...ent first-order condition

$$\phi_i'(x_i^{**}) \leq p_i^{**}, \quad \text{with equality if } x_i^{**} > 0. \tag{11.C.6}$$

... is now viewed as producing a bundle of I goods with a fixed-proportions ... (i.e., the level of production of each personalized good is necessarily the ... Thus, the firm solves

$$\underset{q \geq 0}{\text{Max}} \quad \left(\sum_{i=1}^{I} p_i^{**} q \right) - c(q).$$

... equilibrium level of output q^{**} therefore satisfies the necessary and

sufficient first-order condition

$$\sum_{i=1}^{I} p_i^{**} \leq c'(q^{**}), \quad \text{with equality if } q^{**} > 0. \tag{11.C.7}$$

Together, (11.C.6), (11.C.7), and the market-clearing condition that $x_i^{**} = q^{**}$ for all i imply that

$$\sum_{i=1}^{I} \phi_i'(q^{**}) \leq c'(q^{**}), \quad \text{with equality if } q^{**} > 0. \tag{11.C.8}$$

Comparing (11.C.8) with (11.C.1), we see that the equilibrium level of the public good consumed by each consumer is exactly the efficient level: $q^{**} = q^\circ$.

This type of equilibrium in personalized markets for the public good is known as a *Lindahl equilibrium*, after Lindahl (1919). [See also Milleron (1972) for a further discussion.] To understand why we obtain efficiency, note that once we have defined personalized markets for the public good, each consumer, taking the price in her personalized market as given, fully determines her own level of consumption of the public good; externalities are eliminated.

Yet, despite the attractive properties of Lindahl equilibria, their realism is questionable. Note, first, that the ability to exclude a consumer from use of the public good is essential if this equilibrium concept is to make sense; otherwise a consumer would have no reason to believe that in the absence of making any purchases of the public good she would get to consume none of it.[11] Moreover, even if exclusion is possible, these are markets with only a single agent on the demand side. As a result, price-taking behavior of the sort presumed is unlikely to occur.

The idea that inefficiencies can in principle be corrected by introducing the right kind of markets, encountered here and in Section 11.B, is a very general one. In particular cases, however, this "solution" may or may not be a realistic possibility. We encounter this issue again in our study of multilateral externalities in Section 11.D. As we shall see, these types of externalities often share many of the features of public goods.

11.D Multilateral Externalities

In most cases, externalities are felt and generated by numerous parties. This is particularly true of those externalities, such as industrial pollution, smog caused by automobile use, or congestion, that are widely considered to be "important" policy problems. In this section, we extend our analysis of externalities to these multilateral settings.

An important distinction can be made in the case of multilateral externalities according to whether the externality is *depletable* (or *private*, or *rivalrous*) or *nondepletable* (or *public*, or *nonrivalrous*). Depletable externalities have the feature that experience of the externality by one agent reduces the amount that will be felt by other agents. For example, if the externality takes the form of the dumping of garbage on people's property, if an additional unit of garbage is dumped on one

11. Thus, the possibility of exclusion can be important for efficient supply of the public good, even though the use of an exclusion technology is itself inefficient (a Pareto optimal allocation cannot involve any exclusion).

...of property, that much less is left to be dumped on others.[12] Depletable ...alities therefore share the characteristics of our usual (private) sort of commod... In contrast, air pollution is a nondepletable externality; the amount of air ...on experienced by one agent is not affected by the fact that others are also ...encing it. Nondepletable externalities therefore have the characteristics of ... goods (or bads).

... this section we argue that a decentralized market solution can be expected to ... well for multilateral depletable externalities as long as well-defined and ...able property rights can be created. In contrast, market-based solutions are ... to work in the nondepletable case, in parallel to our conclusions regarding ... goods in Section 11.C.

... shall assume throughout this section that the agents who generate externalities ...inct from those who experience them. This simplification is inessential but ... exposition and facilitates comparison with the previous sections (Exercise ... asks you to consider the general case). For ease of reference, we assume here ... generators of the externality are firms and that those experiencing the ...ty are consumers. We also focus on the special, but central, case in which ...rnality generated by the firms is homogeneous (i.e., consumers are indifferent ... source of the externality). (Exercise 11.D.4 asks you to consider the case in ... the source matters.)

... again adopt a partial equilibrium approach and assume that agents take as ... price vector p of L traded goods. There are J firms that generate the ...ty in the process of production. As discussed in Section 11.B, given price ... we can determine firm j's derived profit function over the level of the ...ty it generates, $h_j \geq 0$, which we denote by $\pi_j(h_j)$. There are also I consumers, ... quasilinear utility functions with respect to a numeraire, traded commodity. ... price vector p, we denote by $\phi_i(\tilde{h}_i)$ consumer i's derived utility function over ...nt of the externality $\tilde{h}_i$ she experiences. We assume that $\pi_j(\cdot)$ and $\phi_i(\cdot)$ are ...erentiable with $\pi_j''(\cdot) < 0$ and $\phi_i''(\cdot) < 0$. To fix ideas, we shall focus on the ...re $\phi_i'(\cdot) < 0$ for all i, so that we are dealing with a negative externality.

...able Externalities

... by examining the case of depletable externalities. As in Section 11.B, it is ... that the level of the (negative) externality is excessive at an unfettered ... equilibrium. Indeed, at any competitive equilibrium, each firm j will wish ... externality-generating activity at the level h_j^* satisfying the condition

$$\pi_j'(h_j^*) \leq 0, \quad \text{with equality if } h_j^* > 0.^{13} \tag{11.D.1}$$

... any Pareto optimal allocation involves the levels $(\tilde{h}_1^\circ, \ldots, \tilde{h}_I^\circ, h_1^\circ, \ldots, h_J^\circ)$

12. A distinction can also be made as to whether a depletable externality is *allocable*. For ... rain is depletable in the sense that the total amount of chemicals put into the air will fall ... but it is not readily allocable because where it falls is determined by weather patterns. ... this section, we take depletable externalities to be allocable. The analytical implications ...cable depletable externalities parallel those of nondepletable ones.

13. ... firms are indifferent about which consumer is affected by their externality. Therefore, ... the fact that $\sum_i \tilde{h}_i = \sum_j h_j^*$, the particular values of the individual $\tilde{h}_i$'s are indeterminate.

that solve[14]

$$\begin{aligned} \underset{\substack{(h_1,\dots,h_J)\geq 0 \\ (\tilde{h}_1,\dots,\tilde{h}_I)>0}}{\text{Max}} \quad & \sum_{i=1}^{I} \phi_i(\tilde{h}_i) + \sum_{j=1}^{J} \pi_j(h_j) \\ \text{s.t.} \quad & \sum_{j=1}^{J} h_j = \sum_{i=1}^{I} \tilde{h}_i. \end{aligned} \tag{11.D.2}$$

The constraint in (11.D.2) reflects the depletability of the externality: If $\tilde{h}_i$ is increased by one unit, there is one unit less of the externality that needs to be experienced by others. Letting μ be the multiplier on this constraint, the necessary and sufficient first-order conditions to problem (11.D.2) are

$$\phi_i'(\tilde{h}_i^\circ) \leq \mu, \quad \text{with equality if } \tilde{h}_i^\circ > 0,\ i = 1, \dots, I, \tag{11.D.3}$$

and

$$\mu \leq -\pi_j'(h_j^\circ), \quad \text{with equality if } h_j^\circ > 0,\ j = 1, \dots, J. \tag{11.D.4}$$

Conditions (11.D.3) and (11.D.4), along with the constraint in problem (11.D.2), characterize the optimal levels of externality generation and consumption. Note that they exactly parallel the efficiency conditions for a private good derived in Chapter 10, conditions (10.D.3) to (10.D.5), where we interpret $-\pi_j'(\cdot)$ as firm j's marginal cost of producing more of the externality. If well-defined and enforceable property rights can be specified over the externality, and if I and J are large numbers so that price taking is a reasonable hypothesis, then by analogy with the analysis of competitive markets for private goods in Chapter 10, a market for the externality can be expected to lead to the optimal levels of externality production and consumption in the depletable case.

Nondepletable Externalities

We now move to the case in which the externality is nondepletable. To be specific, assume that the level of the externality experienced by *each* consumer is $\sum_j h_j$, the total amount of the externality produced by the firms.

In an unfettered competitive equilibrium, each firm j's externality generation h_j^* again satisfies condition (11.D.1). In contrast, any Pareto optimal allocation involves externality generation levels $(h_1^\circ, \dots, h_J^\circ)$ that solve

$$\underset{(h_1,\dots,h_J)>0}{\text{Max}} \quad \sum_{i=1}^{I} \phi_i(\textstyle\sum_j h_j) + \sum_{j=1}^{J} \pi_j(h_j). \tag{11.D.5}$$

This problem has necessary and sufficient first-order conditions for each firm j's optimal level of externality generation, h_j°, of

$$\sum_{i=1}^{I} \phi_i'(\textstyle\sum_j h_j^\circ) \leq -\pi_j'(h_j^\circ), \quad \text{with equality if } h_j^\circ > 0. \tag{11.D.6}$$

14. The objective function in (11.D.2) amounts to the usual difference between benefits and costs arising in the aggregate surplus measure. Note, to this effect, that $\pi_j(\cdot)$ can be viewed as firm j's cost function for producing the externality.

...tion (11.D.6) is exactly analogous to the optimality condition for a public good, ...tion (11.C.1), where $-\pi_j'(\cdot)$ is firm j's marginal cost of externality production.[15]

By analogy with our discussion of private provision of public goods in Section 11.C, the introduction of a standard sort of market for the externality will *not* lead ... as it did in the bilateral case of Section 11.B, to an optimal outcome. The ...der problem reappears, and the equilibrium level of the (negative) externality ... exceed its optimal level. Instead, in the case of a multilateral nondepletable ...ality, a market-based solution would require personalized markets for the ...ality, as in the Lindahl equilibrium concept. However, all the problems with ...hl equilibrium discussed in Section 11.C will similarly afflict these markets. As ... purely market-based solutions, personalized or not, are unlikely to work in ... case of a depletable externality.[16]

... contrast, given adequate information (a strong assumption!), the government ... achieve optimality using quotas or taxes. With quotas, the government simply ... an upper bound on each firm j's level of externality generation equal to its ... level h_j°. On the other hand, as in Section 11.B, optimality-restoring taxes ... each firm with the marginal social cost of their externality. Here the optimal tax ...tical for each firm and is equal to $t_h = -\sum_i \phi_i'(\sum_j h_j^\circ)$ per unit of the externality ...rated. Given this tax, each firm j solves

$$\underset{h_j \geq 0}{\text{Max}} \quad \pi_j(h_j) - t_h h_j,$$

...ch has the necessary and sufficient first-order condition

$$\pi_j'(h_j) \leq t_h, \quad \text{with equality if } h_j > 0.$$

... $t_h = -\sum_i \phi_i'(\sum_j h_j^\circ)$, firm j's optimal choice is $h_j = h_j^\circ$.

A ...rtial market-based approach that can achieve optimality with a nondepletable ...ral externality involves specification of a quota on the *total* level of the externality ... ...ribution of that number of *tradeable externality permits* (each permit grants a firm ... to generate one unit of the externality). Suppose that $h^\circ = \sum_j h_j^\circ$ permits are given ...ms, with firm j receiving $\bar{h}_j$ of them. Let p_h^* denote the equilibrium price of these ... Then each firm j's demand for permits, h_j, solves $\text{Max}_{h_j \geq 0}\,(\pi_j(h_j) + p_h^*(\bar{h}_j - h_j))$ and ... the necessary and sufficient first-order condition $\pi_j'(h_j) \leq p_h^*$, with equality if $h_j > 0$. ...n, market clearing in the permits market requires that $\sum_j h_j = h^\circ$. The competitive ...m in the market for permits then has price $p_h^* = -\sum_i \phi_i'(h^\circ)$ and each firm j using ... and so yields an optimal allocation. The advantage of this scheme relative to a ...ta method arises when the government has limited information about the $\pi_j(\cdot)$ ... and cannot tell which particular firms can efficiently bear the burden of externality ..., although it has enough information, perhaps of a statistical sort, to allow the ...ation of the optimal aggregate level of the externality, h°.

---

15. ...call that the single firm's cost function $c(\cdot)$ in Section 11.C could be viewed as the aggregate ...n of J separate profit-maximizing firms. Were we to explicitly model these J firms in ... 11.C, the optimality conditions for public good production would take exactly the form in ...6) with $c_j'(h_j^\circ)$ replacing $-\pi_j'(h_j^\circ)$.

16. ... public nature of the externality leads to similar free-rider problems in any bargaining ... (See Exercise 11.D.6 for an illustration.)

11.E Private Information and Second-Best Solutions

In practice, the degree to which an agent is affected by an externality or benefits from a public good will often be known only to her. The presence of *privately held* (or *asymmetrically held*) information can confound both centralized (e.g., quotas and taxes) and decentralized (e.g., bargaining) attempts to achieve optimality. In this section, we provide an introduction to these issues, focusing for the sake of specificity on the case of a bilateral externality such as that studied in Section 11.B. Following the convention adopted in Section 11.D, we shall assume here that the externality-generating agent is a firm and the affected agent is a consumer. (For a more general treatment of some of the topics covered in this section, see Chapter 23.)

Suppose, then, that we can write the consumer's derived utility function from externality level h (see Section 11.B for more on this construction) as $\phi(h, \eta)$, where $\eta \in \mathbb{R}$ is a parameter, to be called the consumer's *type*, that affects the consumer's costs from the externality. Similarly, we let $\pi(h, \theta)$ denote the firm's derived profit given its type $\theta \in \mathbb{R}$. The actual values of θ and η are *privately observed*: Only the consumer knows her type η, and only the firm observes its type θ. The ex ante likelihoods (probability distributions) of various values of θ and η are, however, publicly known. For convenience, we assume that θ and η are independently distributed. As previously, we assume that $\pi(h, \theta)$ and $\phi(h, \eta)$ are strictly concave in h for any given values of θ and η.

Decentralized Bargaining

Consider the decentralized approach to the externality problem first. In general, bargaining in the presence of bilateral asymmetric information will *not* lead to an efficient level of the externality. To see this, consider again the case in which the consumer has the right to an externality-free environment, and the simple bargaining process in which the consumer makes a take-it-or-leave-it offer to the firm. For simplicity, we assume that there are only two possible levels of the externality, 0 and $\bar{h} > 0$, and we focus on the case of a negative externality in which externality level $\bar{h}$, relative to the level 0, is detrimental for the consumer and beneficial for the firm (the analysis is readily applied to the case of a positive externality).

It is convenient to define $b(\theta) = \pi(\bar{h}, \theta) - \pi(0, \theta) > 0$ as the measure of the firm's benefit from the externality-generating activity when its type is θ. Similarly, we let $c(\eta) = \phi(0, \eta) - \phi(\bar{h}, \eta) > 0$ give the consumer's cost from externality level $\bar{h}$. In this simplified setting, the only aspects of the consumer's and firm's types that matter are the values of b and c that these types generate. Hence, we can focus directly on the various possible values of b and c that the two agents might have. Denote by $G(b)$ and $F(c)$ the distribution functions of these two variables induced by the underlying probability distributions of θ and η (note that, given the independence of θ and η, b and c are independent). For simplicity, we assume that these distributions have associated density functions $g(b)$ and $f(c)$, with $g(b) > 0$ and $f(c) > 0$ for all $b > 0$ and $c > 0$.

Since the consumer has the right to an externality-free environment, in the absence of any agreement with the firm she will always insist that the firm set $h = 0$ (recall that $c > 0$). However, in any arrangement that guarantees Pareto optimal outcomes for all values of b and c, the firm should be allowed to set $h = \bar{h}$ whenever $b > c$.

Now consider the amount that the consumer will demand from the firm when ... cost is c in exchange for permission to engage in the externality-generating ...ty. Since the firm knows that the consumer will insist on $h = 0$ if there is no ...ement, the firm will agree to pay the amount T if and only if $b \geq T$. Hence, the ...umer knows that if she demands a payment of T, the probability that the firm ... accept her offer equals the probability that $b \geq T$; that is, it is equal to $1 - G(T)$. ... her cost $c > 0$ (and assuming risk neutrality), the consumer optimally chooses ... value of T she demands to solve

$$\underset{T}{\text{Max}} \quad (1 - G(T))(T - c). \tag{11.E.1}$$

...jective function of problem (11.E.1) is the probability that the firm accepts ...mand, multiplied by the net gain to the consumer when this happens $(T - c)$. ... our assumptions, the objective function in (11.E.1) is strictly positive for all ... and equal to zero when $T = c$. Therefore, the solution, say T_c^*, is such that ... But this implies that this bargaining process must result in a strictly positive ...ility of an inefficient outcome, since whenever the firm's benefit b satisfies ... T_c^*, the firm will reject the consumer's offer, resulting in an externality level of ... even though optimality requires that $h = \bar{h}$.[17,18]

...s and Taxes

... decentralized bargaining will involve inefficiencies in the presence of privately ...ormation, so too will the use of quotas and taxes. Moreover, as originally ... Weitzman (1974), the presence of asymmetrically held information causes ... policy instruments to no longer be perfect substitutes for one another, as ... were in the model of Section 11.B.[19]

...egin, note that given θ and η, the aggregate surplus resulting from externality ... (we return to a continuum of possible externality levels here) is $\phi(h, \eta) + \pi(h, \theta)$. ... the externality level that maximizes aggregate surplus depends in general on ...zed values of (θ, η). We denote this optimal value by the function $h(\theta, \eta)$. ... 11.E.1 depicts this optimum value for two different pairs of parameters, (θ', η') ...).

...pose, first, that a quota level of $\hat{h}$ is fixed. The firm will then choose the level ... externality to solve

$$\begin{aligned} \underset{h \geq 0}{\text{Max}} \quad & \pi(h, \theta) \\ \text{s.t.} \quad & h \leq \hat{h}. \end{aligned}$$

... its optimal choice by $h^q(\hat{h}, \theta)$. The typical effect of the quota will be to make

17. Note the similarity between problem (11.E.1) and the monopolist's problem studied in ... 2.B. Here the consumer's inability to discriminate among firms of different types leads her ... offer to be one that yields an inefficient outcome.

18. We could, of course, also consider the outcomes from other, perhaps more elaborate, ...ng procedures. In Chapter 23, however, we shall study a result due to Myerson and ...aite (1983) that implies that *no* bargaining procedure can lead to an efficient outcome ... values of b and c in this setting.

19. The discussion that follows also has implications for the relative advantages of quantity- ... price-based control mechanisms in organizations.

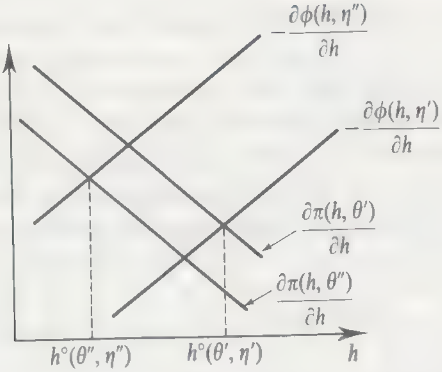


Figure
The
max
exter
two
para
$(\theta',$

the actual level of the externality much less sensitive to the values of θ and η than is required by optimality. The firm's externality level will be completely insensitive to η. Moreover, if the level of the quota $\hat{h}$ is such that $\partial\pi(\hat{h},\theta)/\partial h > 0$ for all θ, we will have $h^q(\hat{h},\theta) = \hat{h}$ for every θ. The loss in aggregate surplus arising under the quota for types (θ,η) is given by

$$\phi(h^q(\hat{h},\theta),\eta) + \pi(h^q(\hat{h},\theta),\theta) - \phi(h^\circ(\theta,\eta),\eta) - \pi(h^\circ(\theta,\eta),\theta)$$
$$= \int_{h^\circ(\theta,\eta)}^{h^q(\hat{h},\theta)} \left(\frac{\partial\pi(h,\theta)}{\partial h} + \frac{\partial\phi(h,\eta)}{\partial h}\right) dh.$$

This loss is represented by the shaded region in Figure 11.E.2 for a case in which the quota is set equal to $\hat{h} = h^\circ(\bar{\theta},\bar{\eta})$, the externality level that maximizes social surplus when θ and η each take their mean values, $\bar{\theta}$ and $\bar{\eta}$ [the dashed lines in the figure are the graphs of $\partial\pi(h,\bar{\theta})/\partial h$ and $-\partial\phi(h,\bar{\eta})/\partial h$ and the solid lines are the graphs of $\partial\pi(h,\theta)/\partial h$ and $-\partial\phi(h,\eta)/\partial h$; note that in the case depicted, the firm wishes to produce the externality up to the allowed quota $\hat{h}$].

Consider next the use of a tax on the firm of t units of the numeraire per unit of the externality. For any given value of θ, the firm will then choose the level of externality to solve

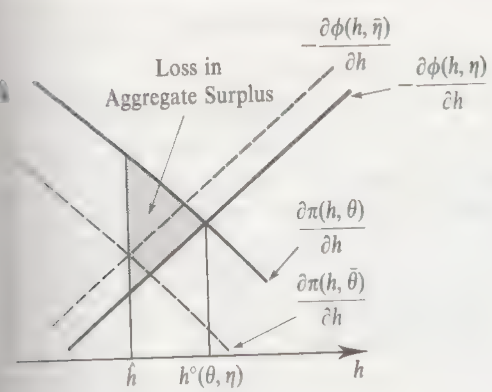
$$\underset{h \geq 0}{\text{Max}} \quad \pi(h,\theta) - th.$$

Denote its optimal choice by $h^t(t,\theta)$. The loss in aggregate surplus from the tax relative to the optimal outcome for types (θ,η) is therefore given by

$$\phi(h^t(t,\theta),\eta) + \pi(h^t(t,\theta),\theta) - \phi(h^\circ(\theta,\eta),\eta) - \pi(h^\circ(\theta,\eta),\theta)$$
$$= \int_{h^\circ(\theta,\eta)}^{h^t(t,\theta)} \left(\frac{\partial\pi(h,\theta)}{\partial h} + \frac{\partial\phi(h,\eta)}{\partial h}\right) dh.$$

Its value is depicted by the shaded region in Figure 11.E.3 for the same situation as in Figure 11.E.2, but now assuming that a tax is set at $t = -\partial\phi(h^\circ(\bar{\theta},\bar{\eta}),\bar{\eta})/\partial h$, the value that results in the maximization of aggregate surplus when $(\theta,\eta) = (\bar{\theta},\bar{\eta})$. Note that under a tax, as under a quota, the level of the externality is responsive to changes in the marginal benefits of the firm but not to changes in the marginal costs of the consumer.

Which of these instruments, quota or tax, performs better? The answer is that it depends. Imagine, for example, that η is a constant, say equal to $\bar{\eta}$. Then, for θ such that the benefits of the externality's use to the firm are high, a quota will typically miss the optimal externality level by not allowing the externality to increase above the quota level. On the other hand, because a fixed tax rate t does not reflect any



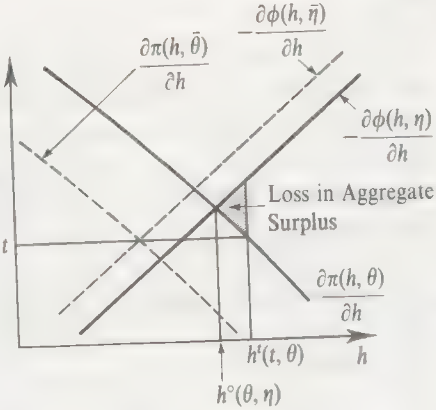


Figure 11.E.2 (left) The loss in aggregate surplus under a quota for types (θ, η).

Figure 11.E.3 (right) The loss in aggregate surplus under a tax for types (θ, η).

...creasing marginal costs of the externality to the consumer at higher externality ...ds, for such values of θ the tax may result in excess production of the externality. ...itively, when the optimal externality level varies little with θ, we expect a ...to be better. Figure 11.E.4(a), for example, depicts a case in which the marginal ... the consumer of the externality is zero up to some point h^* and infinite ...ter. In this case, by setting a quota of $\hat{h} = h^*$, we can maximize aggregate ... for any value of (θ, η), but no tax can accomplish this. A tax would have to ... high to guarantee that with probability one the externality level fixed by the ... not larger than h^*. But if so, the resulting externality level would be too low ... of the time.

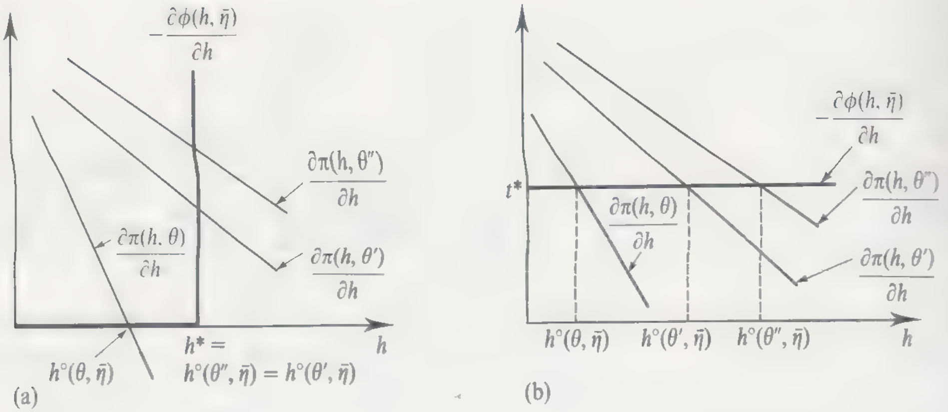
In contrast, in Figure 11.E.4(b) we depict a case in which the marginal cost to ... consumer of the externality is independent of the level of h. In this case, a tax ... to this marginal cost ($t = t^*$) achieves the surplus-maximizing externality level ... all (θ, η), but no quota can do so.

...e take the expected value of aggregate surplus as our welfare measure, we ...e see from these two examples that either policy instrument may be preferable, ...ding on the circumstances.[20] (Exercise 11.E.1 asks you to provide a full ... for a linear-quadratic example.) Note also that the bargaining procedure we ...iscussed will not result in optimality in *either* case depicted in Figure 11.E.4.[21] ...e have here two cases in which either a quota or a tax performs better than ...icular decentralized outcome.[22]

...In Chapter 13, we discuss in greater detail some of the issues that arise in making welfare ...ons in settings with privately held information. There we shall justify the maximization of ... aggregate surplus in this partial equilibrium setting as a requirement of a notion of ex ...eto optimality for the two agents. See also the discussion in Section 23.F.

...trictly speaking, our previous discussion of bargaining assumed only two possible levels of ...nality, while here we have a continuum of levels. This difference is not important. The ...cy of the bargaining procedure previously studied would hold in this continuous environ... as well.

...We should emphasize that in these two examples other bargaining procedures will perform ...an the procedure involving a take-it-or-leave-it offer by the consumer. For example, if a ...eave offer is made by the firm, then full optimality results in *both* of these cases because ... of the consumer is known with certainty. The conclusion of our discussion is therefore a ...e one: With asymmetric information, it is difficult to make very general assertions about the ...e performance of centralized versus decentralized approaches.



In Exercise 11.E.2, you are asked to extend the analysis just given to a case with two firms ($j = 1, 2$) generating an externality, where the two firms are identical except possibly for their realized levels of θ_j. The exercise illustrates the importance of the degree of correlation between the θ_j's for the relative performance of quotas versus taxes. In comparing a uniform quota policy versus a uniform tax policy ("uniform" here means that the two firms face the same quota or tax rate), the less correlated the shocks across the firms, the better the tax looks. The reason is not difficult to discern. With imperfect correlation, a uniform tax has a benefit that is not achieved with a uniform quota: It allows for the individual levels of externalities generated to be responsive to the realized values of the θ_j's. Indeed, with a uniform tax, the production of the total amount of externality generated is always efficiently distributed across the two firms.

The presence of multiple generators of an externality also raises the possibility that a market for tradeable emissions permits could be created, as discussed at the end of Section 11.D. This simple addition to the quota policy can potentially eliminate the inefficient distribution of externality generation across different generators that is often a feature of a quota policy. In particular, suppose that instead of simply giving each firm a quota level, we now give them tradeable externality permits entitling them to generate the same number of units of the externality as in the quota. Suppose also that each firm would always fully use its quota if no trade was possible. Then trade must result in *at least* as large a value of aggregate surplus as the simple quota scheme for any realization of the firms' and consumer's types, because we still get the same total level of emissions and we can never get a trade between firms that lowers aggregate profits.[23] Of course, the same bargaining problems that we studied above can prevent a fully efficient distribution of externality generation from arising; but if the firms know each others' values of θ_j or are numerous enough to act competitively in the market for these rights, then we can expect a distribution of the total externality generation that is efficient across generators. In fact, in the case where the statistical distribution of costs among the firms is known but the particular realizations for individual firms are not known, this type of policy can achieve a fully optimal outcome.

23. Note, however, that the assumption that the externalities generated by the different firms are perfect substitutes to the consumer is crucial to this conclusion. If this is not true, then the reallocation of externality generation can reduce aggregate surplus by lowering the well-being of the agents affected by the externality.

More General Policy Mechanisms

The tax and quota schemes considered above are, as we have seen, completely unresponsive to changes in the marginal costs of the externality to the affected agent (the consumer in this case). It is natural to wonder whether any other sorts of schemes can do better, perhaps by making the level of the externality responsive to the consumer's costs. The problem in doing so is that these benefits and costs are unobservable, and the parties involved may not have incentives to reveal them truthfully if asked. For example, suppose that the government simply asks the consumer and the firm to report their benefits and costs from the externality and then enforces whatever appears to be the optimal outcome based on these reports. In this case, the consumer will have an incentive to exaggerate her costs when asked in order to prevent the firm from being allowed to generate the externality. The question, then, is how to design mechanisms that control these incentives for misreporting and, as a consequence, enable the government to achieve an efficient outcome. This problem is studied in a very general form in Chapter 23; here we restrict ourselves to a brief examination of one well-known scheme.

Return to the case in which there are only two possible levels of the externality, 0 and $\bar{h}$. Can we design a scheme that achieves the optimal level of externality generation for every realization of b (the firm's benefit from the externality) and c (the consumer's cost)? We now verify that the answer is "yes."

Imagine the government setting up the following *revelation mechanism*: The firm and the consumer are each asked to report their values of b and c, respectively. Let $\hat{b}$ and $\hat{c}$ denote these announcements. For each possible pair of announcements $(\hat{b}, \hat{c})$, the government sets an allowed level of the externality as well as a tax or subsidy payment for each of the two agents. Suppose, in particular, that the government announces that it will set the allowed externality level h to maximize aggregate surplus given the announcements. That is, $h = \bar{h}$ if and only if $\hat{b} > \hat{c}$. In addition, if externality generation is allowed (i.e., if $h = \bar{h}$), the government will tax the firm an amount equal to $\hat{c}$ and will subsidize the consumer with a payment equal to $\hat{b}$. That is, if the firm wants to generate the externality (which it indicates by reporting a large value of b), it is asked to pay the externality's cost as declared by the consumer; and if the consumer allows the externality (by reporting a low value of c) she receives a payment equal to the externality's benefit as declared by the firm.

In fact, *under this scheme both the firm and the consumer will tell the truth*, so that the optimal level of externality generation will, indeed, result for every possible (b, c) pair. To see this, consider the consumer's optimal announcement when her cost level is c. If the firm announces some $\hat{b} > c$, then the consumer prefers to have the externality-generating activity allowed (she does $\hat{b} - c$ better than if it is prevented). Hence, her optimal announcement satisfies $\hat{c} < \hat{b}$; moreover, because any such announcement will give her the same payoff, she might as well announce the truth, that is, $\hat{c} = c < \hat{b}$. On the other hand, if the firm announces $\hat{b} \leq c$, the consumer prefers to have the externality level set to zero. Hence, she would like announce $\hat{c} \geq \hat{b}$; and again, because any of these announcements will give her the same payoff, she might as well announce the truth, that is, $\hat{c} = c \geq \hat{b}$. Thus, whatever the firm's announcement, truth-telling is an optimal strategy for the consumer. (Formally, telling the truth is a *weakly dominant strategy* for the consumer in the sense studied

in Section 8.B. In fact, it is the consumer's *only* weakly dominant strategy; see Exercise 11.E.3.) A parallel analysis yields the same conclusion for the firm.

Exercise 11.E.4: Show that in the tax-subsidy part of the mechanism above we could add, without affecting the mechanism's truth-telling or optimality properties, an additional payment to each agent that depends in an arbitrary way on the other agent's announcement.

The scheme we have described here is an example of the *Groves–Clarke mechanism* [due to Groves (1973) and Clarke (1971); see also Section 23.C] and was originally proposed as a mechanism for deciding whether to carry out public good projects. Some examples for the public goods context are contained in the exercises at the end of the chapter.

The Groves–Clarke mechanism has two very attractive features: it implements the optimal level of the externality for every (b, c) pair, and it induces truth-telling in a very strong (i.e., dominant strategy) sense. But the mechanism has some unattractive features as well. In particular, it does not result in a balanced budget for the government: The government has a deficit equal to $(b - c)$ whenever $b > c$. We could use the flexibility offered by Exercise 11.E.4 to eliminate this deficit for all possible (b, c), but then we would necessarily create a budget surplus and therefore a Pareto inefficient outcome for some values of (b, c) (not all units of the numeraire will be left in the hands of the firm or the consumer).

In fact, this problem is unavoidable with this type of mechanism: If we want to preserve the properties that, for every (b, c), truth-telling is a dominant strategy and the optimal level of externality is implemented, then we generally cannot achieve budget balance for every (b, c). In Chapter 23 we discuss this issue in greater detail and also consider other mechanisms that can, under certain circumstances, get around the problem. (See also Exercise 11.E.5 for an analysis in which budget balance is required only on average.)

APPENDIX A: NONCONVEXITIES AND THE THEORY OF EXTERNALITIES

Throughout this chapter, we have maintained the assumption that preferences and production sets are convex, leading the derived utility and profit functions we have considered to be concave. With these assumptions, all the decision problems we have studied have been well behaved; they had unique solutions (or, more generally, convex-valued solutions) that varied continously with the underlying parameters of the problems (e.g., the prices of the L traded commodities or the price of the externality if a market existed for it). Yet, this is not a completely innocent assumption. In this appendix, we present some simple examples designed to illustrate that externalities may themselves generate nonconvexities, and we comment on some of the implications of this fact.

We consider here a bilateral externality situation involving two firms. We suppose that firm 1 may engage in an externality-generating activity that affects firm 2's production. The level of externality generated by firm 1 is denoted by h, and firm j's profits conditional on the production of externality level h are $\pi_j(h)$ for $j = 1, 2$. It is perfectly natural to assume that $\pi_1(\cdot)$ is concave: The level h could, for example,

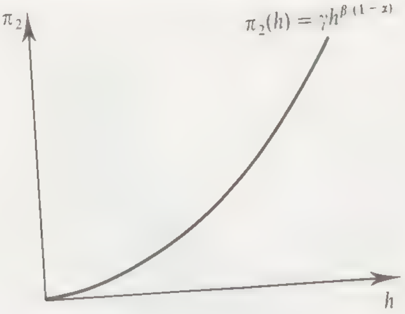


Figure 11.AA.1
The derived profit function of firm 2 (the externality recipient) in Example 11.AA.1 when $\alpha + \beta > 1$.

equal to firm 1's output.[24] As Examples 11.AA.1 and 11.AA.2 illustrate, however, s may not be true of firm 2's profit function.

ample 11.AA.1: *Positive Externalities as a Source of Increasing Returns.* Suppose firm 2 produces an output whose price is 1, using an input whose price, for plicity, we also take to equal 1. Firm 2's production function is $q = h^{\beta} z^{\alpha}$, where $\in [0, 1]$. Thus, the externality is a positive one.[25] Note that, for fixed h, the blem of firm 2 is concave and perfectly well behaved. Given a level of h, the ximized profits of firm 2 can be calculated to be $\pi_2(h) = \gamma h^{\beta/(1-\alpha)}$, where $\gamma > 0$ is onstant. In Figure 11.AA.1, we represent $\pi_2(h)$ for $\beta > 1 - \alpha$. We see there that 2's derived profit function is *not* concave in h; in fact, it is convex. This reflects fact that if we think of the externality h as an input to firm 2's production , then firm 2's overall production function exhibits increasing returns to scale e $\alpha + \beta > 1$. ■

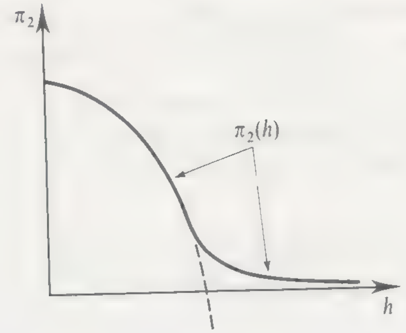
mple 11.AA.2: *Negative Externalities as a Source of Nonconvexities.* In Example A.1, the nonconvexity in firm 2's production set, and the resulting failure of avity in its derived profit function, were caused by a positive externality. In this ple the failure of concavity of firm 2's derived profit function is the result of a ive externality.

Suppose, in particular, that $\pi_2'(h) \leq 0$ for all h, with strict inequality for some h, that firm 2 has the option of shutting down when experiencing externality level receiving profits of zero.[26] In this case, the function $\pi_2(\cdot)$ can *never* be concave

24. Note also that we may well have $\pi_1(h) < 0$ for some levels $h \geq 0$ because $\pi_1(h)$ is firm 1's mized profit *conditional* on producing externality level h (and so shutting down is not possible

25. More generally, we could think that there is an industry composed of many firms and that ernality is produced and felt by all firms in the industry (e.g., h could be an index, correlated output, of accumulated know-how in the industry). Externalities were first studied by Marshall this context. See also Chipman (1970) and Romer (1986).

26. In the more typical case of a multilateral externality, the ability of affected parties to shut in this manner often depends on whether the externality is depletable. In the case of a pletable externality, such as air pollution, affected firms can always shut down and receive ofits. In contrast, in the case of a depletable externality (such as garbage), where $\pi_j(h)$ reflects 's profits when it individually absorbs h units of the externality, the absorption of the externality self require the use of some inputs (e.g., land to absorb garbage). Indeed, were this not the a depletable externality, the externality could always be absorbed in a manner that creates al costs by allocating all of the externality to a firm that shuts down.



over all $h \in [0, \infty)$, a point originally noted by Starrett (1972). The reason can be seen in Figure 11.AA.2: If $\pi_2(\cdot)$ were a strictly decreasing concave function, then it would have to become negative at some level of h (see the dashed curve), but $\pi_2(\cdot)$ must be nonnegative if firm 2 can always choose to shut down. ■

The failure of $\pi_2(\cdot)$ to be concave can create problems for both centralized and decentralized solutions to the externality problem. For example, if property rights over the externality are defined and a market for the externality is introduced in either Example 11.AA.1 or Example 11.AA.2, a competitive equilibrium may fail to exist (even assuming that the two agents act as price takers). Firm 2's objective function will not be concave, and so its optimal demand may fail to be well defined and continuous (recall our discussion in Section 10.C of the equilibrium existence problems caused by nonconvexities in firms' cost functions).

In contrast, taxes and quotas can, in principle, still implement the optimal outcome despite the failure of firm 2's profit function to be concave because their use depends only on the profit function of the externality generator (here, firm 1) being well behaved. In practice, however, nonconvexities in firm 2's profit function may create problems for these centralized solutions as well. Example 11.AA.3 illustrates this point.

Example 11.AA.3: *Externalities as a Source of Multiple Local Social Optima.* It is, in principle, true that if the decision problem of the generator of an externality is concave, then the optimum can be sustained by means of quotas or taxes. But if $\pi_2(\cdot)$ is not concave, then the aggregate surplus function $\pi_1(h) + \pi_2(h)$ may not be concave and, as a result, the first-order conditions for aggregate surplus maximization may suffice only for determining local optima. In fact, as emphasized by Baumol and Oates (1988), the nonconvexities created by externalities may easily generate situations with multiple local social optima, so that identifying a global optimum may be a formidable task.

Suppose, for example, that the profit functions of the two firms are

$$\pi_1(h) = \begin{cases} h & \text{for } h \le 1 \\ 1 & \text{for } h > 1 \end{cases}$$

and

$$\pi_2(h) = \begin{cases} 2(1-h)^2 & \text{for } h \le 1 \\ 0 & \text{for } h > 1. \end{cases}$$

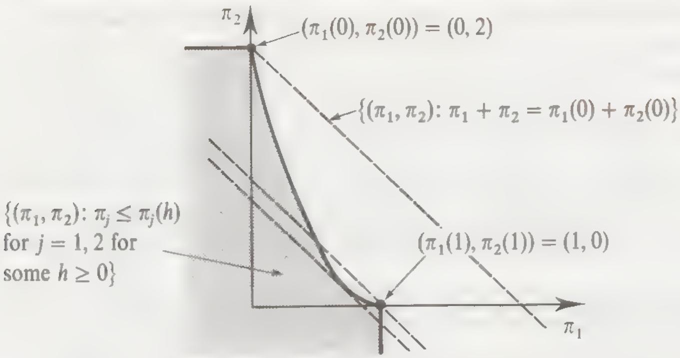


Figure 11.AA.3 The set of possible profit pairs (π_1, π_2) in Example 11.AA.3 exhibits multiple local maxima of aggregate surplus $\pi_1(h) + \pi_2(h)$.

function $\pi_2(\cdot)$ is not concave, something that the two previous examples have n us can easily happen with externalities. The profit levels for the two firms that attainable for different levels of h are depicted in Figure 11.AA.3 by the shaded $\pi_1, \pi_2)$: $\pi_j \le \pi_j(h)$ for $j = 1, 2$ for some $h > 0\}$ (note that this definition for free disposal of profits). The social optimum has $h = 0$ (joint profits are equal to 2), in which case firm 2 is able to operate in an environment free from ternality. This can be implemented by setting a tax rate on firm 1 of $t > 1$ per the externality. But note that the outcome $h = 1$ (implemented by setting a te on firm 1 of $t = 0$) is a local social optimum: As we decrease h, it is not until that we get an aggregate surplus level higher than that at $h = 1$. Hence, this outcome satisfies both the first-order and second-order conditions for the ization of aggregate surplus (e.g., at this point, the marginal benefits of the ality exactly equal its marginal costs), and it will be easy for a social planner be misled into thinking that she is at a welfare maximum. ■

ERENCES

K. J. (1969). The organization of economic activity: Issues pertinent to the choice of market versus non-market allocation. In *Collected Papers of K. J. Arrow*, Vol. 2. Cambridge, Mass.: Harvard University Press, 1983.

W. J. (1964). External economies and second-order optimality conditions. *American Economic Review* **54**: 368–72.

W. J., and W. E. Oates. (1988). *The Theory of Environmental Policy*, 2nd. ed. New York: Cambridge University Press.

J. S. (1970). External economies of scale and competitive equilibrium. *Quarterly Journal of Economics* **84**: 347–85.

E. H. (1971). Multipart pricing of public goods. *Public Choice* **11**: 17–33.

R. (1960). The problem of social cost. *Journal of Law and Economics* **1**: 1–44.

T. (1973). Incentives in teams. *Econometrica* **41**: 617–31.

strom, B., and J. Tirole. (1989). The theory of the firm. In *Handbook of Industrial Organization*, edited by R. Schmalensee and R. D. Willig. Amsterdam: North-Holland.

J.-J. (1988). *Fundamentals of Public Economics*. Cambridge, Mass.: MIT Press.

E. (1919). *Die Gerechtigkeit der Besteuring*. Lund: Gleerup. [English translation: Just taxation—a positive solution. In *Classics in the Theory of Public Finance*, edited by R. A. Musgrave and A. T. Peacock. London: Macmillan, 1958.]

A. (1920). *Principles of Economics*. London: Macmillan.

Meade, J. (1952). External economies and diseconomies in a competitive situation. *Economic Journal* **62**: 54–67.

Milleron, J.-C. (1972). Theory of value with public goods: A survey article. *Journal of Economic Theory* **5**: 419–77.

Myerson, R., and M. Satterthwaite. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory* **29**: 265–81.

Pigou, A. C. (1932). *The Economics of Welfare.* London: Macmillan.

Romer, P. (1986). Increasing returns and long-run growth. *Journal of Political Economy* **94**: 1002–36.

Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics* **36**: 387–89.

Samuelson, P. A. (1955). Diagrammatic exposition of a pure theory of public expenditure. *Review of Economics and Statistics* **37**: 350–56.

Starrett, D. A. (1972). Fundamental non-convexities in the theory of externalities. *Journal of Economic Theory* **4**: 180–99.

Viner, J. (1931). Cost curves and supply curves. *Zeitschrift für Nationalökonomie* **111**: 23–46.

Weitzman, M. (1974). Prices vs. quantities. *Review of Economic Studies* **41**: 477–91.

EXERCISES

11.B.1[B] (M. Weitzman) On Farmer Jones' farm, only honey is produced. There are two ways to make honey: with and without bees. A bucket full of artificial honey, absolutely indistinguishable from the real thing, is made out of 1 gallon of maple syrup with one unit of labor. If the same honey is made the old-fashioned way (with bees), k total units of labor are required (including bee-keeping) and b bees are required per bucket. Either way, Farmer Jones has the capacity to produce up to H buckets of honey on his farm.

The neighboring farm, belonging to Smith, produces apples. If bees are present, less labor is needed because bees pollinate the blossoms instead of workers doing it. For this reason, c bees replace one worker in the task of pollinating. Up to A bushels of apples can be grown on Smith's farm.

Suppose that the market wage rate is w, bees cost p_b per bee, and maple syrup costs p_m per gallon. If each farmer produces her maximal output at the cheapest cost to her (assume the output prices they face make maximal production efficient), is the resulting outcome efficient? How does the answer depend on k, b, c, w, p_b, and p_m? Give an intuitive explanation of your result. Up to how much would Smith be willing to bribe Jones to produce honey with bees? What would happen to efficiency if both farms belonged to the same owner? How could the government achieve efficient production through taxes?

11.B.2[C] Consider the two-consumer externality problem studied in Section 11.B, but now assume that consumer 2's derived utility function over the externality level h and her wealth available for commodity purchases w_2 takes the form $\phi_2(h, w_2)$. Assume that $\phi_2(h, w_2)$ is a twice-differentiable, strictly quasiconcave function with $\partial\phi_2(h, w_2)/\partial w_2 > 0$ and, for simplicity, that we have a positive externality so that $\partial\phi_2(h, w_2)/\partial h > 0$.

(a) Set up the Pareto optimality problem as one of choosing h and a wealth transfer T to maximize consumer 1's welfare subject to giving consumer 2 a utility level of at least $\bar{u}_2$. Derive the (necessary and sufficient) first-order condition characterizing the optimal levels of h and T, say h° and T°.

(b) Imagine that consumer 1 could purchase h on an externality market. Let p_h be the price per unit, and let $h(p_h, w_2)$ be consumer 2's demand function for h. Express the wealth effect $\partial h(p_h, w_2)/\partial w_2$ in terms of first-order and second-order partial derivatives of consumer 2's utility function.

(c) Derive the comparative statics change in the Pareto optimal level of the externality h° (for a given $\bar{u}_2$) with respect to a differential increase $dw_2 > 0$ in consumer 2's wealth. Show that if consumer 2's demand for the externality, derived in (b), is normal at price $\hat{p}_h = [\partial\phi_2(h^\circ, w_2 - T^\circ)/\partial h]/[\partial\phi_2(h^\circ, w_2 - T^\circ)/\partial w_2]$ and wealth level $\bar{w}_2 = w_2 - T^\circ$ [i.e., if $\partial h(\hat{p}_h, \bar{w}_2)/\partial w_2 > 0$], then a marginal increase in consumer 2's wealth w_2 causes the Pareto optimal level of the externality h° to increase. (Similarly, in the case of a negative externality, if consumer 2's demand for reductions in the externality is a normal good, then when consumer 2 becomes wealthier, the Pareto optimal level of the externality declines.)

11.B.3[B] Consider the optimal Pigouvian tax identified in Section 11.B for the two-consumer externality problem studied there. What happens if, given this tax, the two consumers are able to bargain with each other? Will the efficient level of the externality still result? What about with the optimal quota?

11.B.4[B] Consider again the two-consumer externality problem studied in Section 11.B. Suppose that consumer 2 can take some action, say $e \in \mathbb{R}$, that affects the degree to which she is affected by the externality, so that we now write her derived utility function as $\phi_2(h, e) + w_2$. To fix ideas, let h be a negative externality, and suppose that $\partial^2\phi_2(h, e)/\partial h \partial e > 0$, so that increases in e reduce the negative effect of the externality on the margin. Suppose that both h and e can in principle be taxed or subsidized. Should e be taxed or subsidized in the optimal tax scheme? Why or why not?

11.B.5[B] Suppose that at fixed input prices of $\bar{w}$ a firm produces output with the differentiable and strictly convex cost function $c(q, h)$, where $q \geq 0$ is its output level (whose price is $p > 0$) and h is the level of a negative externality generated by the firm. The externality affects a single consumer, whose derived utility function takes the form $\phi(h) + w$. The actions of the firm and consumer do not affect any market prices.

(a) Derive the first-order condition for the firm's choice of q and h.

(b) Derive the first-order conditions characterizing the Pareto optimal levels of q and h.

(c) Suppose that the government taxes the firm's output level. Show that this cannot restore efficiency. Show that a direct tax on the externality *can* restore efficiency.

(d) Show, however, that in the limiting case where h is necessarily produced in fixed proportions with q, so that $h(q) = \alpha q$ for some $\alpha > 0$, a tax on the firm's output *can* restore efficiency. What is the efficiency-restoring tax?

11.C.1[A] Consider the model discussed in Section 11.C, in which I consumers privately purchase a public good. Identify per-unit subsidies $s_1, \ldots, s_I$, such that when each consumer i faces subsidy rate s_i, the total level of the public good provided is optimal.

11.C.2[A] Consider the model discussed in Section 11.C, in which I consumers privately purchase a public good. Show that a per-unit subsidy on the firm's output (paid to the firm) can also restore efficiency.

11.C.3[C] Reconsider the Ramsey tax problem from Exercise 10.E.3, but now suppose that the government can also provide a public good x_0 that can be produced from good 1 at cost $c(x_0)$. However, the government must still balance its budget (including any expenditures on the public good). Consumer i's utility function now takes the form $x_{1i} + \sum_{\ell=2}^{L} \phi_{\ell i}(x_{\ell i}, x_0)$. Derive and interpret the conditions characterizing the optimal commodity taxes and the optimal level of the public good. How do the two problems of Ramsey taxation and provision of the public good interact?

11.D.1[B] (M. Weitzman) First-year graduate students are a hard-working group. Consider a typical class of I students. Suppose that each student i puts in h_i hours of work on her classes. This effort involves a disutility of $h_i^2/2$. Her benefits depend on how she performs relative to her peers and take the form $\phi(h_i/\bar{h})$ for all i, where $\bar{h} = (1/I)\sum_i h_i$ is the average number of hours put in by all students in the class and $\phi(\cdot)$ is a differentiable concave function, with

$\phi'(\cdot) > 0$ and $\lim_{h \to 0} \phi'(h) = \infty$. Characterize the symmetric (Nash) equilibrium. Compare it with the Pareto optimal symmetric outcome. Interpret.

11.D.2^B Consider a setting with I consumers. Each consumer i chooses an action $h_i \in \mathbb{R}_+$. Consumer i's derived utility function over her choice of h and the choices of other consumers takes the form $\phi_i(h_i, \sum_i h_i) + w_i$, where $\phi_i(\cdot)$ is strictly concave. Characterize the optimal levels of $h_1, \ldots, h_I$. Compare these with the equilibrium levels. What tax/subsidy scheme induces the optimal outcome?

11.D.3^B Consider an industry composed of $J > 1$ identical firms that act as price takers. The price of their output is p, and the prices of their inputs are unaffected by their actions. Suppose that partial equilibrium analysis is valid and that the aggregate demand for their product is given by the function $x(p)$. The industry is characterized by "learning by doing," in that each firm's total cost of producing a given level of output is declining in the level of total industry output; that is, each firm j has a twice-differentiable cost function of the form $c(q_j, Q)$ for $Q = \sum_j q_j$, where $c(\cdot)$ is strictly increasing in its first argument and strictly decreasing in its second. Letting subscripts denote partial derivatives, assume that $c_q + Jc_Q > 0$ and $(1/n)c_{qq} + 2c_{qQ} + nc_{QQ} > 0$ for $n = 1$ and J. Compare the equilibrium and optimal industry output levels. Interpret. What tax or subsidy restores efficiency?

11.D.4^B Reconsider the nondepletable externality example discussed in Section 11.D, but now assume that the externalities produced by the J firms are not homogeneous. In particular, suppose that if $h_1, \ldots, h_J$ are the firms' externality levels, then consumer i's derived utility is given by $\phi_i(h_1, \ldots, h_J) + w_i$ for each $i = 1, \ldots, I$. Compare the equilibrium and efficient levels of $h_1, \ldots, h_J$. What tax/subsidy scheme can restore efficiency? Under what condition should each firm face the same tax/subsidy rate?

11.D.5^B (*The problem of the commons*) Lake Ec can be freely accessed by fishermen. The cost of sending a boat out on the lake is $r > 0$. When b boats are sent out onto the lake, $f(b)$ fish are caught in total [so each boat catches $f(b)/b$ fish], where $f'(b) > 0$ and $f''(b) < 0$ at all $b \geq 0$. The price of fish is $p > 0$, which is unaffected by the level of the catch from Lake Ec.

(a) Characterize the equilibrium number of boats that are sent out on the lake.

(b) Characterize the optimal number of boats that should be sent out on the lake. Compare this with your answer to **(a)**.

(c) What per-boat fishing tax would restore efficiency?

(d) Suppose that the lake is instead owned by a single individual who can choose how many boats to send out. What level would this owner choose?

11.D.6^B Suppose that there is a piece of land that is affected adversely by an externality produced by a single firm. The firm's derived profit function for the externality is $\pi(h) = \alpha + \beta h - \mu h^2$, where h is the level of the externality and $(\alpha, \beta, \mu) \gg 0$. There are I consumers who farm the land, each owning a fraction $1/I$ of it. The total yield of the land is $\phi(h) = \gamma - \eta h$, where $(\gamma, \eta) \gg 0$. Each of the I consumers then has a derived utility function of $\phi(h)/I + w$.

Bargaining among the consumers and the firm works as follows: Each consumer simultaneously decides whether to be in or out of a bargaining coalition. After this, the bargaining coalition makes the firm a take-it-or-leave-it offer specifying a level of h and a transfer. The firm then accepts or rejects this offer. In the absence of any agreement, the firm can generate any level of the externality it wishes.

(a) Let θ denote the fraction of the I consumers who join the bargaining coalition. Characterize the subgame perfect Nash equilibrium level of θ (for simplicity, treat θ as a continuous variable). Show that when $I = 1$ the optimal level of the externality results, but that when $I > 1$ we have $\theta < 1$ in equilibrium and too much of the externality is generated.

(b) Show that as I increases, the equilibrium level of θ declines. Also show that $\lim_{I \to \infty} \theta = 0$.

11.D.7[C] Individuals can build their houses in one of two neighborhoods, A or B. It costs c_A to build a house in neighborhood A and $c_B < c_A$ to build in neighborhood B. Individuals care about the prestige of the people living in their neighborhood. Individuals have varying levels of prestige, denoted by the parameter θ. Prestige varies between 0 and 1 and is uniformly distributed across the population. The prestige of neighborhood k $(k = A, B)$ is a function of the average value of θ in that neighborhood, denoted by $\bar{\theta}_k$. If individual i has prestige parameter θ and builds her house in neighborhood k, her derived utility net of building costs is $(1 + \theta)(1 + \bar{\theta}_k) - c_k$. Thus, individuals with more prestige value a prestigous neighborhood more. Assume that c_A and c_B are less than 1 and that $c_A - c_B \in (\frac{1}{2}, 1)$.

(a) Show that in any building-choice equilibrium (technically, the Nash equilibrium of the simultaneous-move game in which individuals simultaneously choose where to build their house) both neighborhoods must be occupied.

(b) Show that in any equilibrium in which the prestige levels of the two neighborhoods differ, every resident of neighborhood A must have at least as high a prestige level as every resident of neighborhood B; that is, there is a cutoff level of θ, say $\hat{\theta}$, such that all types $\theta \geq \hat{\theta}$ build in neighborhood A and all $\theta < \hat{\theta}$ build in neighborhood B. Characterize this cutoff level.

(c) Show that in any equilibrium of the type identified in (b), a Pareto improvement can be achieved by altering the cutoff value of θ slightly.

11.E.1[B] Consider the setting studied in Section 11.E, and suppose that $\partial\pi(h, \theta)/\partial h = \beta - bh + \theta$ and $\partial\phi(h, \eta)/\partial h = \gamma - ch + \eta$, where θ and η are random variables with $E[\theta] = E[\eta] = E[\theta\eta] = 0$, $(\beta, b, c) \gg 0$, and $\gamma < 0$. Denote $E[\theta^2] = \sigma_\theta^2$ and $E[\eta^2] = \sigma_\eta^2$.

(a) Identify the best quota $\hat{h}^*$ for a planner who wants to maximize the expected value of aggregate surplus. (Assume the firm must produce an amount exactly equal to the quota.)

(b) Identify the best tax t^* for this same planner.

(c) Compare the two instruments: Which is better and when?

11.E.2[C] Extend the model in Exercise 11.E.1 to the case of two producers. Now let $\partial\pi_i(h_i, \theta_i)/\partial h = \beta - bh_i + \theta_i$ for $i = 1, 2$. Let $\sigma_{12} = E[\theta_1\theta_2]$. Calculate and compare the optimal quotas and taxes. How does the choice depend on σ_{12}?

11.E.3[B] Show that truth-telling is the consumer's only weakly dominant strategy in the (Groves–Clarke) revelation mechanism studied in Section 11.E.

11.E.4[A] In text.

11.E.5[B] Suppose that the government is considering building a public project. The cost is K. There are I individuals indexed by i. Individual i's privately known benefit from the project is b_i. The government's objective is to maximize the expected value of aggregate surplus. Derive the extension of the Groves–Clarke mechanism discussed in Section 11.E for this case. Can you construct your scheme so that the government balances its budget on average (over all realizations of the b_i's)?

11.E.6[B] Extend Exercise 11.E.5 to the case in which there are N possible projects, $n = 1, \ldots, N$, with individual i deriving a (privately known) benefit of $b_i(n)$ from project n.

11.E.7[B] Suppose that in the model of Section 11.E the consumer's type η takes only one possible value, $\bar{\eta}$. We have seen in the text that in this case neither a quota nor a tax will maximize aggregate surplus for all realizations of θ when the derived utility function $\phi(h, \bar{\eta})$ for the consumer has $\partial\phi(h, \bar{\eta})/\partial h \in (0, -\infty)$. Show, however, that a variable tax per unit in which the total tax collected from the firm is $\phi(h, \bar{\eta})$ when the level of the externality is h will maximize aggregate surplus for all values of θ for *any* derived utility function $\phi(h, \bar{\eta})$.